

**IMPROVING THE ANALYSIS OF FOREIGN AFFAIRS:
EVALUATING STRUCTURED ANALYTIC TECHNIQUES**

By

Stephen J. Coulthart

B.A. in Political Science, State University of New York at Oswego, 2006

M.A. in International Relations and Diplomacy, Seton Hall University, 2010

M.P.A in Public Administration, Seton Hall University, 2010

Submitted to the Graduate Faculty of the Graduate School of Public
and International Affairs in partial fulfillment of the requirements for
the degree of PhD in Public and International Affairs

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH
GRADUATE SCHOOL OF PUBLIC AND INTERNATIONAL AFFAIRS

This dissertation was presented
By

Stephen Coulthart

It was defended on
June 4, 2015
and approved by

Shawn Bird, PhD, Division Chief, U.S. Department of State-Bureau of Intelligence
and Research

Michael Kenney, PhD, Associate Professor, Graduate School of Public and
International Affairs, University of Pittsburgh

Phil Williams, PhD, Professor, Graduate School of Public and International Affairs,
University of Pittsburgh

Dissertation Chair: William N. Dunn, PhD, Professor, Graduate School of Public and
International Affairs, University of Pittsburgh

Copyright © by Stephen J. Coulthart
2015

IMPROVING ANALYSIS OF FOREIGN AFFAIRS: EVALUATING STRUCTURED ANALYTIC TECHNIQUES

Stephen Coulthart, M.P.A./M.P.A

Research suggests that foreign affairs analysis is weak – even the best analysts are accurate less than 35 percent of the time (Tetlock 2005). To compensate for analytic weaknesses, some have called for the use of structured analytic techniques, that is, formalized intelligence analysis methods. This imperative was enshrined in the Intelligence Reform and Terrorism Prevention Act (2004), which mandates that analysts use these techniques. This research investigates how the techniques have been applied in the U.S. intelligence community (IC) while making a modest attempt to evaluate 12 core techniques.

The investigation of how the techniques are applied is based on semi-structured interviews with 5 intelligence experts and a survey of 80 analysts at an IC agency, along with follow-up interviews with 15 analysts. Interestingly, 1 in 3 analysts reported never using the techniques. Two factors were related to the use of the techniques: analytic training ($p=0.001$, Cramer's $V=0.41$) and the perception of their value ($p=.049$, Cramér's $V=0.23$). There was not a statistically significant relation between the time pressure under which analysts work and their use of the techniques ($p=0.74$).

Questions about the effectiveness of the techniques were answered in part by employing a “systematic review,” a novel methodology for synthesizing a large body of research. A random sample of more than 2,000 studies, suggests that there is moderate to

strong evidence affirming the efficacy of using three techniques: Analysis of Competing Hypotheses, Brainstorming, and Devil's Advocacy. There were three main findings: face-to-face collaboration decreases creativity, evidence weighting appears to be more important than seeking disconfirming evidence, and conflict tends to improve the quality of analysis. This research also employed an experiment with 21 graduate intelligence studies students, which confirmed the first two findings of the systematic review.

The findings of the dissertation represent a contribution to "evidence-based intelligence analysis," the systematic effort to develop a robust evidence-base linking the use of specific analytic techniques to the improvement of analysis in foreign affairs. Future research might build on the evidence-base presented here to improve intelligence analysis, one of the most important areas of judgment in foreign affairs.

TABLE OF CONTENTS

| | | |
|-------|--|----|
| 1.0 | CHAPTER 1: ADDRESSING THE LIMITS OF FOREIGN AFFAIRS ANALYSIS | 1 |
| 1.1 | THE LIMITATIONS OF EXPERTISE IN FOREIGN AFFAIRS ANALYSIS.... | 2 |
| 1.2 | SPANNING THE DIVIDE: IMPROVING DECISION AND ANALYSIS IN FOREIGN AFFAIRS | 4 |
| 1.2.1 | The Theory-Application Gap- Evaluating the Intelligence Reform Act | 6 |
| 1.3 | DEVELOPING A THEORY OF STRUCTURED ANALYTIC TECHNIQUES | 10 |
| 1.3.1 | A Theory of Structured Analytic Techniques | 13 |
| 1.4 | ADOPTING EVIDENCE-BASED INTELLIGENCE ANALYSIS TO ANSWER THE RESEARCH QUESTIONS..... | 17 |
| 1.4.1 | Research Question 1: The Implementation of Structured Analytic Techniques | 18 |
| 1.4.2 | Research Question 2: The Effectiveness of Structured Analytic Techniques | 20 |
| 1.5 | AT THE FRONTIER OF FOREIGN AFFAIRS ANALYSIS | 24 |
| 2.0 | CHAPTER 2: ASSESSING THE QUALITY OF INTELLIGENCE ANALYSIS: A SYNTHESIS AND REFORMULATION | 26 |
| 2.1 | WHAT IS ‘QUALITY’ INTELLIGENCE ANALYSIS AND WHY IT’S SO HARD TO PRODUCE? | 28 |

| | | |
|-------|--|----|
| 2.1.1 | Defining Intelligence Analysis Quality | 29 |
| 2.1.2 | The Diminishing Returns of Expert Judgment..... | 34 |
| 2.2 | REFRAMING INTELLIGENCE ANALYSIS AS AN “INEXACT SCIENCE” | 38 |
| 2.2.1 | Intelligence Analysis as an “Inexact Science” | 39 |
| 2.2.2 | A Justification for Structured Analytic Techniques: System 1 and 2 Thinking | 44 |
| 2.2.3 | A New Theory of Structured Analytic Techniques | 46 |
| 2.3 | IMPLEMENTING STRUCTURED ANALYTIC TECHNIQUES IN THE IC | 51 |
| 2.3.1 | The Answer is in the Head, not the Numbers..... | 51 |
| 2.3.2 | New Era, New Analysis | 54 |
| 2.3.3 | Intelligence Reform and Terrorism Prevention Act of 2004 and Large- Scale Implementation..... | 57 |
| 2.3.4 | Factors Affecting the Implementation of Structured Analytic Techniques | 59 |
| 2.4 | SUMMARIZING THE ARGUMENT | 62 |
| 3.0 | CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY | 65 |
| 3.1 | FIELD STUDY: IMPLEMENTING STRUCTURED ANALYTIC TECHNIQUES | 66 |
| 3.1.1 | Semi-structured Interviews with Intelligence Analysis Experts..... | 67 |
| 3.1.2 | Survey and Interviews at the Bureau of Intelligence and Research.. | 68 |
| 3.2 | EVALUATING STRUCTURED ANALYTIC TECHNIQUES: A SYSTEMATIC REVIEW AND FIELD EXPERIMENT | 76 |

| | | |
|-------|---|-----|
| 3.2.1 | A Systematic Review of Structured Analytic Techniques..... | 77 |
| 3.2.2 | An Experiment of ACH and Indicators | 84 |
| 3.3 | CONCLUSION..... | 94 |
| 4.0 | CHAPTER 4: WHO USES STRUCTURED ANALYTIC TECHNIQUES AND WHY?: A SURVEY IN THE INTELLIGENCE COMMUNITY | 96 |
| 4.1 | FRAMING THE STUDY: TESTING THE VARIABLES AT STATE DEPARTMENT INR | 97 |
| 4.1.1 | Study Hypotheses..... | 98 |
| 4.1.2 | State Department’s Bureau of Intelligence and Research | 99 |
| 4.1.3 | Gauging the Use of Structured Analytic Techniques at INR | 101 |
| 4.2 | STUDY HYPOTHESIS 1: ANALYST TRAINING..... | 102 |
| 4.3 | PERCEPTIONS OF THE TECHNIQUES, TIME PRESSURE AND DEMOGRAPHICS | 106 |
| 4.3.1 | Perception of Structured Analytic Techniques..... | 107 |
| 4.3.2 | Time Pressure | 110 |
| 4.3.3 | Demographics: Age | 112 |
| 4.4 | SUMMARY OF RESULTS AND CONCLUSIONS | 115 |
| 5.0 | CHAPTER 5: “WHAT WORKS?” A SYSTEMATIC REVIEW OF STRUCTURED ANALYTIC TECHNIQUES..... | 117 |
| 5.1 | OVERVIEW: “WHAT WORKS?” | 120 |
| 5.1.1 | Selecting and Describing the Evaluative Studies | 121 |
| 5.1.2 | Assessing the Credibility of Evidence..... | 123 |

| | | |
|-------|---|-----|
| 5.2 | UNDER WHAT CONDITIONS ARE THE TECHNIQUES EFFECTIVE? | 127 |
| 5.2.1 | Analysis of Competing Hypotheses Studies..... | 128 |
| 5.2.2 | Analysis of Competing Hypotheses Discussion | 130 |
| 5.2.3 | Brainstorming Studies | 132 |
| 5.2.4 | Brainstorming Discussion..... | 136 |
| 5.2.5 | Devil's Advocacy Studies..... | 137 |
| 5.2.6 | Devil's Advocacy Discussion | 142 |
| 5.3 | SUMMARY OF FINDINGS | 143 |
| 6.0 | CHAPTER 6: AN EXPERIMENT OF ANALYSIS OF COMPETING HYPOTHESES..... | 145 |
| 6.1 | FRAMING THE STUDY | 147 |
| 6.1.1 | Task Background and Procedures | 147 |
| 6.1.2 | Study Hypotheses..... | 149 |
| 6.2 | TESTING THE STUDY HYPOTHESES..... | 151 |
| 6.2.1 | Study Hypotheses 1 and 2: Collaboration and Multiple Rival Hypotheses | 151 |
| 6.2.2 | Study Hypothesis 3: Hypothesis Disconfirmation and Accuracy | 156 |
| 6.2.3 | Study Hypothesis 4: Interaction effect of the techniques and reasoning styles | 160 |
| 6.3 | DISCUSSION OF RESULTS..... | 164 |

| | | |
|-------|---|-----|
| 6.3.1 | Face-to-Face Collaboration is a Limiter of Creativity and Next Steps | 164 |
| 6.3.2 | Beyond Confirmation Bias: Evidence Weighting..... | 166 |
| 6.3.3 | Cognitive Reasoning Style and Structured Analytic Techniques ... | 167 |
| 6.4 | SUMMARY AND CONCLUSIONS | 170 |
| 7.0 | CHAPTER 7: EVIDENCE-BASED PRINCIPLES FOR IMPROVING INTELLIGENCE ANALYSIS..... | 172 |
| 7.1 | EVIDENCE-BASED PRINCIPLES FOR INTELLIGENCE ANALYSIS | 174 |
| 7.1.1 | Training and the Value of Evidence | 174 |
| 7.1.2 | Avoid Face-to-Face Collaboration | 176 |
| 7.1.3 | Focus on Evidence Weighting | 178 |
| 7.1.4 | Harness Conflict Carefully | 180 |
| 7.2 | SEEKING TO BE “APPROXIMATELY RIGHT”: SUMMARY AND POLICY IMPLICATIONS | 183 |
| 8.0 | METHODOLOGICAL APPENDIX A: SURVEY AND INTERVIEW PROCEDURES | 186 |
| 9.0 | METHODOLOGICAL APPENDIX B: SYSTEMATIC REVIEW PROCEDURES | 191 |
| 10.0 | METHODOLOGICAL APPENDIX C: EXPERIMENT PROCEDURES..... | 216 |
| 11.0 | BIBLIOGRPAHY..... | 221 |

LIST OF TABLES

Table 1.1: “Core” Structured Analytic Techniques

Table 1.2 Selected Rigor Attributes and Description

Table 2.1: Truncated and Reproduced Rigor Attributes from Zelik et. al. (2010)

Table 2.2: Structured Analytic Techniques

Table 3.1: Research Questions

Table 3.2: Selected Interview Questions and Variables

Table 3.3: Survey Variables and Questions

Table 3.4 Follow-up Interviews

Table 3.5: Excluded Studies

Table 3.6: Truncated Indicators Grid

LIST OF FIGURES

Figure 3.1: Formula for Estimating Sample Size (adapted from Dillman et al. 2011)

Figure 3.2: Calculation of Appropriate Sample Size

Figure 3.3: The Maryland Scientific Methods Scale

Figure 3.4: Intelligence Task Complexity

Figure 3.5 An Example ACH Matrix

Figure 3.6: Snippet of Narrative with Hypotheses

Figure 3.7: Example of Distribution of Hypotheses

Figure 4.1 Use of Structured Analytic Techniques at INR

Figure 4.2: Perception of Techniques (Rigor)

Figure 4.3: Analysts' Self-Reported Average Analytic Product at INR

Figure 4.4: Age Cohorts at INR

Figure 5.1: Techniques Included in this Study

Figure 5.2: Number of Evaluative Studies

Figure 5.3: Modified Maryland Scientific Methods Scale

Figure 5.4: What Works?: A Display of the Overall Results

Figure 6.1: Indicators group's hypotheses before using the technique

Figure 6.2: Indicators group's hypotheses after using the technique

Figure 6.3: ACH group's hypotheses before using the technique

Figure 6.4: ACH group's hypotheses after using the technique

Figure 6.5: Between Groups Comparison

Figure 6.6: Pretest and Posttest of Indicators

Figure 6.7: Pretest and Posttest of ACH

Figure 6.8: Foxes hypotheses before using a technique

Figure 6.9: Hedgehogs hypotheses before using a technique

Figure 6.10: Foxes hypotheses after using a technique

Figure 6.11: Hedgehogs hypotheses after using a technique

Figure 6.12 Cumulative Frequency of Hypotheses

PREFACE

For Beatriz:

*“Caminante, son tus huellas el camino, y nada más; caminante, no hay camino, se hace camino
al andar...”*

We did it.

1.0 CHAPTER 1: ADDRESSING THE LIMITS OF FOREIGN AFFAIRS ANALYSIS

“Good judgment” wrote Mark Twain “is the result of experience and experience the result of bad judgment.” Decades of psychological research suggest Twain was on to something: experts excel in domains where they can try, and try again, failing, and then learning from their mistakes, to develop expert judgment (Kahneman and Klein, 2009). In these domains experts are able to learn by trial and error and develop accurate, internalized models of a problem situation upon which they can draw to make judgments (Simon, 1965). Consider the chess master who plays game after game, learning from each failure and, in the process, stores tens of thousands of distinct move patterns in his mind (Chase and Simon, 1973), or the fireman who, after responding to thousands of structure fires, intuitively “knows” from his experience that a backdraft is on the other side of the door. In each case the decision maker has had enough opportunities to develop expert judgment on thousands of chances to make ‘bad judgments.’¹

¹ Similar findings exist at the organizational-scale. For example Petroski (2008) argues that in order to improve performance, engineering should look to past failures rather than successes. For an extended discussion with examples, see: Petroski, Henry. *Success through failure: The paradox of design*. Princeton University Press, 2006.

1.1 THE LIMITATIONS OF EXPERTISE IN FOREIGN AFFAIRS ANALYSIS

Analysts of foreign affairs, who study issues of war and peace, face a gloomier situation: unlike the chess master or firefighter, the foreign affairs analyst has little or no feedback because the events of interest rarely occur, or in the case of nuclear war, hopefully never occur.² Even when foreign affairs analysts get feedback, it is all too easy for political and ideological reasons for the analysts to ignore or write off their inaccurate judgments (Tetlock 2005). Take for example, the foreign policy “hawk” who might view the collapse of the USSR not as a moment to recalibrate future predictions about aggressors, but as an unlikely “one-off” occurrence. Complicating matters even further, the foreign affairs analysts lack valid indicators (Kaheman and Klein, 2009). While the firefighter has many indicators to tell him what is or will happen, such as a weak support beam suggesting imminent collapse, the foreign affairs decision maker has almost none. For example, a column of armor massing on the Ukrainian border can mean different things: Is a Russian invasion is imminent? Or is the armored column just a sign of Russian resolve? In short, unlike other professionals with more valid indicators and repetitive and clear feedback, the foreign affairs expert has few ways to develop expert intuitive judgment.

² There are exceptions in political forecasting, most notably the statistical forecasting by Nate Silver in U.S. elections. Statistical analysis “works” in these cases for reasons similar to those that enable the chess master to acquire precise intuitive judgment: ample and reliable feedback, in the form of survey data, are used to draw inferences, rather than expert judgment. For an extended discussion of this issue, see: Jay Ulfelder “Why the World Can’t Have a Nate Silver,” *Foreign Policy*, November, 2012.

Researchers have long known that foreign affairs analysts and decision makers struggle to make valid judgments, and in the process sometimes commit errors of reasoning, such as underestimating or overestimating the aggressiveness of an adversary. For example, Jervis (1976) pointed out that leaders often see what they expect and/or want to see, and in the process act in unexpected, and dangerous ways. Janis (1972) explored the limitations of expert judgment in foreign affairs at the group-level with his influential work on “groupthink,” a term that has since moved into the vernacular for undue group consensus. However, it was not until Philip Tetlock’s seminal work *Expert Political Judgment* (2005) that the full extent of judgmental deficiency in foreign affairs was systematically laid bare. Over the course of several years Tetlock asked 280 experts to make judgments about foreign events. While Tetlock found his experts outperformed undergraduate students, they “were not superior to untrained readers of newspapers in their ability to make accurate long-term forecasts of political events” (Kahneman and Klein, 2009, p. 520). The experts’ ability to easily outperform undergraduates but inability to significantly top observers suggests that there is a diminishing return expertise. Further, even the *highest performing* experts achieved an accuracy rate only slightly greater than chance alone. Or, in other words, the forecasts of this elite group were, on average, only slightly better than the flip of a coin. It is here, at the frontier of expert competence that the research problem emerges: what can be done to improve foreign affairs judgments?

1.2 SPANNING THE DIVIDE: IMPROVING DECISION AND ANALYSIS IN FOREIGN AFFAIRS

Despite the recognition among scholars that foreign affairs judgment is limited, there is little work in the international relations literature on how to improve it.³ Most of the literature that does address this problem dates back to the 1970s and focuses on improving executive decision making. For example, George (1972) developed “multiple advocacy,” a method⁴ designed to promote diversity of perspectives using an appointed arbiter to control discussion in executive decision making. Another example is Devil’s Advocacy, a method that uses a designated individual or group to take an unpopular position in order to “combat the problems posed by an excessive tendency” toward “consensus-seeking behavior” – Janis’ groupthink (George, 1975, pp. 286–287). Outside of the international relations literature there is a much larger body of research on techniques to improve judgment in environments as complex as foreign affairs. For

³ Stephen Walt provides an excellent explanation on why research on conducting analysis of foreign affairs is limited: “I think it is partly because scholars in international relations have tended to focus on grand theory (realism, liberalism, constructivism, etc.), or on trying to identify recurring laws or tendencies between states or other groups...In other words, most scholars stand apart from the policy process and treat international affairs as something to be studied from a safe distance.” Stephen Walt, “Policy analysis in global affairs: What should my students read?” *ForeignPolicy.com*, November 2011, available at:

http://www.foreignpolicy.com/posts/2011/11/22/policy_analysis_in_global_affairs_what_should_my_students_read

⁴ Kaplan (1973) defines methods in three ways 1) broad methods that involve broad procedures (e.g. induction); 2) mid-range methods sufficiently general to be common to all sciences that have somewhat specific procedures (e.g. forming concepts, using research designs); 3) and short-range methods with specific procedures and purposes (e.g. factor analysis). This research uses the term technique and method interchangeably to refer to short-range methods.

example, a technique very similar to Devil's Advocacy was used to improve the U.S. Census Bureau's planning process (Mitroff *et al.* 1977).

Despite the promising literature that sets forth new methods of foreign affairs analysis, there is little research that explicitly tests their efficacy, although there are some notable exceptions outside the foreign affairs community and international relations literature (for example, see Armstrong 2006). The efficacy question is important from a theoretical standpoint as researchers attempt to develop and test theories on foreign policy decision making and in a practical sense because of the important decisions that are made based on these methods. Consider the empirical evidence for the efficacy of the Alternative Futures Analysis, a technique used to identify a set of important drivers that can lead to various scenarios that, once reported to the organization's leadership can be used to prevent surprise (Schwartz 1991, p. 3-4). Supporters point to examples of effectiveness of scenarios, specifically its use by Royal Dutch Shell in the 1970s. This example suggests because Shell used scenarios it was prepared for rising tensions in the Arab world after the Yom Kippur War. Tetlock (2005, p. 192) is less convinced and argues that because users of the technique write several scenarios, they virtually ensure that at least one will resemble the outcome. Simply stated, it is not possible to know if a technique is effective without rigorous empirical tests, tests which we lack for most methods. One rebuttal to Tetlock is that the techniques are not designed to forecast

specific outcomes, but still, a valid question is whether the scenarios are plausible and internally consistent.⁵

The question of efficacy of foreign affairs methods is of paramount importance because these methods could address the deficiencies of expert judgment and improve decisions that drive high stakes decisions. To situation this dissertation, the inquiry focuses on examining perhaps the largest producer of foreign affairs analysis the world has ever known, the U.S. intelligence community (IC).

1.2.1 The Theory-Application Gap- Evaluating the Intelligence Reform Act

Similar to analysts in other domains, such as policy and business analysis, intelligence analysts apply their expertise to weight data and compare, contrast, and evaluate information concerning important events (Johnston, 2005, p. 3). This process requires that analysts sort through “enormous volumes of data and combine seemingly unrelated events to construct an accurate interpretation of a situation and make predictions about complex, dynamic events” (Pirolli 2005). Intelligence analysts produce intelligence “products,” documents and presentations of different types. These products can vary greatly with some focusing on providing warning to decision maker of a potential threat to providing descriptive research about a particular country or leader. Intelligence

⁵ Thomas Barnett argues that the National Intelligence Council’s *Global Trends 2030* report lacks of internally “consistent logic throughout each of the worlds presented.” Available at: <http://nation.time.com/2012/12/21/just-how-intelligent-is-the-national-intelligence-councils-global-trends-2030/>

analysts also work in high-stakes environments where intelligence errors can occur, factual inaccuracies in analysis (Johnston 2005, p. 6). The avoidance of error cannot be overstated as even slight errors--if reproduced and replicated at the organizational level--can potentially lead to multi-billion dollar intelligence and policy failures.

One such example where intelligence error contributed, at least in part, to an intelligence failure is in the lead-up to the Iraq War. In the National Intelligence Estimate, the IC's authoritative assessment of Iraq's weapons of mass destruction program, all the IC agencies but State Department's Bureau of Intelligence and Research⁶ determined Iraq had continued its programs (National Intelligence Council 2002). After the invasion turned up no weapons of mass destruction, the analytic practices of the IC came under tight scrutiny as panels, committees, and commissions initiated investigations. In a conclusion typical of these probes, one congressional panel co-chairman argued that the IC estimate on Iraq's nuclear program was "inconsistent" and based on "tortured presumptions" (Stout 2005). The resounding message of the investigation was that U.S. intelligence analysis needed nothing less than an analytic transformation, and in 2004 the Intelligence Reform and Terrorism Prevention Act (henceforth, the "Intelligence Reform Act") laid out a legal framework to accomplish exactly that.

⁶ The State Department's Bureau for Intelligence and Research (INR) concluded: "Saddam continues to want nuclear weapons and that available evidence indicates that Baghdad is pursuing at least a limited effort to maintain and acquire nuclear weapons-related capabilities. The activities we have detected do not, however, add up to a compelling case that Iraq is currently pursuing what INR would consider to be an integrated and comprehensive approach to acquire nuclear weapons." National Intelligence Council, "Iraq's Continuing Programs for Weapons of Mass Destruction" (2002), <http://fas.org/irp/cia/product/iraq-wmd.html>

There have been many prior efforts to reform intelligence analysis. What sets the Intelligence Reform Act apart from most other IC reforms is that it is a reform of day-to-day analytic practices to address what Betts (2007) terms an “enemy” of intelligence: the inherent limits of human cognition in intelligence analysis. As a result of the focus on improving the process of analysis, the Intelligence Reform Act and the initiatives it set in motion are sometimes referred to as the “analytic transformation” or “analytic reform movement” (Immerman 2011; Fingar 2011). The added emphasis on analysis, or analytic tradecraft as it is called in the IC, is in contrast to previous reforms that focused on improving information sharing through greater organizational collaboration, restructuring, or a combination of both. In terms of analytic practice, where the Intelligence Reform Act arguably makes the biggest impact is the requirement that analysts be trained in and use methods called structured analytic techniques (section 1017, sub-section A).

While the Intelligence Reform Act mandated structured analytic techniques, their implementation has been gaining momentum in the IC for the last three decades. Beginning in the late 1970s, the IC experimented with formalizing analysis by exploring new analytical methods leading to a focus on techniques designed to avoid certain cognitive biases (Heuer 1978). In the 1990s, changing international threats, a quicker intelligence cycle, and data overload, created more interest in implementing “alternative

analysis techniques”⁷ as the techniques were known at that time. Reformers argued that the techniques could improve analysis by encouraging analysts to consider multiple rival hypotheses and challenge their assumptions. After the Intelligence Reform Act the impetus to implement the techniques grew and reformers sought to import methods from other disciplines including policy analysis, political science, and business. Consequently, there are literally hundreds of possible methods that fall under the rubric of “structured analytic techniques.” To reduce this number to a workable set this research focuses on the twelve “core” techniques identified in the U.S. Government’s *Analytic Tradecraft Primer* (2009) (see Table 1.1 below).

Table 1.1: “Core” Structured Analytic Techniques

| |
|--------------------------------------|
| Key Assumptions Check |
| Quality of Information Check |
| Indicators of Signpost/ Change |
| Analysis of Competing Hypotheses |
| Devil’s Advocacy |
| Team A/Team B |
| High-Impact/Low-Probability Analysis |
| “What If?” Analysis |
| Brainstorming |
| Outside-In Thinking |
| Red Team Analysis |
| Alternative Futures Analysis |

As of 2011, more than 4,000 analysts have received training in these twelve techniques, representing a quarter of the IC’s analytic workforce (Defense Intelligence Agency, 2011). At the same time, this growth is expected to accelerate as more

⁷ Alternative analysis techniques were renamed structured analytic techniques to encourage analysts to think of them not as alternatives but mainstay of “good” analysis (Heuer and Pherson, 2010, p. 9).

intelligence agencies mandate training programs (Heuer and Pherson, 2010, p. 343) and more intelligence analysis education programs train the next generation of analysts. A recent analysis of these education programs found at least 13 that have been founded since 2001 (Crosston and Coulthart, forthcoming).

Notwithstanding the push for training, there are serious barriers for reformers to implement the techniques. For example, ethnographies of the IC suggest that analysts are reluctant to utilize techniques because of the increasing current reporting requirements that emerged in the 1990s and accelerated after September 11th (Johnston 2005, pp. 26-27; Dixon and McNamara 2008). At same time many techniques require a perceived, significant time investment to decompose a problem and externalize the information to paper or a screen (Heuer 1999, p. 86). Understanding how to implement the techniques is important if they can be shown to leverage expertise and improve analysis.

1.3 DEVELOPING A THEORY OF STRUCTURED ANALYTIC TECHNIQUES

Despite the mandate for structured analytic techniques, there is no overall theory explaining why the techniques should improve intelligence analysis. The theoretical justification that does exist is rooted in research from cognitive psychology conducted in the 1970s and 1980s. This research was based on the heuristics and biases research conducted by Daniel Kahneman and Amos Tversky. They found that respondents in lab experiments used mental shortcuts or rules of thumb to make quick judgments, and that

these heuristics led to predictable errors. For example, in a classic study (Tversky and Kahneman 1983), respondents were told about a young woman named Linda:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Respondents were then asked which is more probable: a) Linda is a bank teller, or b) Linda is a bank teller and is active in the feminist movement. Most respondents chose 'b,' and in making this choice, respondents used the representative heuristic. People use this heuristic to quickly estimate probabilities on the basis of previous experience. Since respondents can imagine Linda easier as a feminist bank teller than just a bank teller they disregard a simple fact about probability: the probability of two events occurring in conjunction is always less than one event.

Heuer imported Kahneman and Tversky's findings into intelligence analysis with his classic, *The Psychology of Intelligence Analysis* (1999), which was later the basis for a justification of structured analytic techniques. Heuer argues that the subjects in Kahneman and Tversky's experiments are similar to intelligence analysts, who also use heuristics that increase the likelihood of errors (Heuer 1999). For example, an analyst making a judgment about which foreign terrorist group will attack the U.S. might invoke the availability heuristic, a mental short-cut based on the last or most memorable event. Since al Qaeda is the last perpetrator of a major terrorist attack, the analyst might focus only on this group and therefore commit an error of reasoning, thereby ignoring other known or emerging groups that may attack.

Due to these errors Heuer and Pherson (2014) argue that analysts should use techniques that switch analysts' thinking from quick and intuitive thinking that use heuristics to slow and effortful thinking. Kahneman (2011) refers to these two types of thinking as systems 1 and 2. A common example of system 1 is the type of effortless, intuitive thinking a morning commuter uses to find his way to work, while system 2 thinking is used in difficult reasoning tasks, such as solving a complex math problem. Heuer and Pherson argue that the problem of cognitive biases lays in system 1 because they argue that all biases are the result of fast thinking (p. 4). Therefore, they conclude, the answer lies in using more system 2 thinking, specifically through using structured analytic techniques, which they claim "help identify and overcome the analytic biases inherent in system 1 thinking" (p.5). In short, according to Heuer and Pherson, using structured analytic techniques should engender slow analytic thinking that should reduce cognitive biases.

Even with Heuer and Pherson's notable effort to provide a theoretical justification of structured analytic techniques, there are issues with their explanation. One issue is that Heuer and Pherson focus exclusively on cognitive biases as the main performance standard of the techniques. This is problematic because it narrows the standard for measuring the effectiveness of the techniques. The normative standards in the Linda example above, whether the respondent violated the conjunction fallacy – the belief that two events occurring is more likely than one-- is illustrative. While certainly it matters that analysts not violate rules of formal logic and probability theory, other standards also matter in intelligence analysis, perhaps even more so. For example, while adherence to

probability theory matter to researchers, managers and intelligence consumers appear to view rigor more in line with depth and scope of analysis (Zelik *et al.* 2009).

At the same time, there are serious methodological challenges of proving that a bias violated formal logic and probability theory, as several psychological research studies have demonstrated the difficulty of replicating biases in even highly controlled experiments (Hammond 1996, pp-203-213). This latter point is important for theory development and testing as any justification for structured analytic techniques should show a replicable outcome. In addition, while Heuer and Pherson give the use of system 2 as a rationale for the techniques, this justification does not explain why the techniques might improve analysis because system 2 is as susceptible to biases as system 1 (Kahan *et al.* 2006, pp. 1093-1094). Another problem is that the techniques pull from both systems. For example, the use of Alternative Futures Analysis encourages intuitive, imaginative thinking to conjure possible outcomes and effortful thinking to identify contextual drivers. The issue does not stop here; many other techniques fit this general description, relying on both systems. As a result of these limitations a new causal theory of structured analytic techniques is needed.

1.3.1 A Theory of Structured Analytic Techniques

To address the limitations of Heuer and Pherson's theoretical justification, a causal theory is needed that explains outcomes in terms of prior actions or conditions. In this research the causal theory addresses how the use of structured analytic techniques can improve

the quality of intelligence. Astute readers might notice that the use of “theory” in this context is similar to “method,” however, there is an important distinction. While a method implies a procedure of some kind to reach a desired goal (Kaplan 1973), a theory sets forth a set of propositions to *explain* a particular phenomenon. In this context the theory was formulated to explain how the techniques improve the quality of analysis.

For the purposes of this research the quality of intelligence analysis has two components: 1) it is sufficiently rigorous and 2) accurate. Analysis can be said to be rigorous when it is in-depth, as reflected in intelligence reports. Researchers at the Ohio State University (Miller, Patterson, and Woods, 2006; Zelik, Patterson, and Woods, 2007) identified eight attributes of rigor in intelligence analysis by observing and surveying analysts, thus creating a grounded measure of analytic rigor. These include whether the analyst explored a range of hypotheses, questioned the reliability of sources, among others (see table 1.2, below). The term “sufficient” is also important because sufficiency is dependent on the factors that shape the analytic task: complexity/breadth of topic, data availability, and time (Greitzer 2004; Scholtz and Hewett 2004). Relying on these characteristics it is possible to determine basic “rules of thumb” for sufficient rigor. For example, a complex task, with available data, and a project length of several months, such as a long-term strategic forecast, would require more rigorous analysis than a one day project. It is important to note that this framework subsumes cognitive bias. For example, confirmation bias--the error of seeking to confirm rather than disconfirm one’s beliefs—would be subsumed under “hypothesis exploration” as low hypothesis exploration would be related to confirming a favored hypothesis. In other words, the Ohio State

University scale does not replace cognitive biases but provides a reliable measure grounded in how analysts evaluate rigor.⁸

Table 1.2 Selected Rigor Attributes and Description

| | |
|------------------------|---|
| Hypothesis Exploration | The construction and evaluation of potential explanations for collected data. |
| Information Search | The focused collection of data bearing upon the analysis problem. |
| Information Validation | The critical evaluation of data with respect to the degree of agreement among sources |
| Stance Analysis | The evaluation of collected data to identify the relative positions of sources with respect to the broader contextual setting |

In addition to rigor, analysis is high quality when it is accurate, that is, when there is a high degree of correspondence between the analytic judgment and what subsequently happened in the external world (Tetlock, 2005, p. 10; Hammond 1996). Or in other words, the extent to which an analyst “gets it right.” In intelligence studies, accuracy is sometimes conceptualized as a “batting average” (Betts 2007, p. 187). Yet, assessing accuracy is not as simple as checking the box scores; many tasks in foreign affairs are covert and complex, rendering evaluation of judgment accuracy difficult (e.g. did country X build a nuclear weapon?). For this reason great caution should be exercised in evaluating accuracy.⁹

⁸ Evaluation of the reliability of the Ohio State scale has been positive. In one experiment, intercoder reliability across 12 evaluations was strong between two coders ($\kappa_w = 0.86$) (Zelik *et al.* 2007, p. 11)

⁹ For an excellent discussion of the challenges of evaluating tests of accuracy in foreign affairs, see: Jay Ulfelder, “Jay Ulfelder on the Rigor-Relevance Tradeoff” *The Good Judgment Project*, May, 2014, available at: <https://goodjudgmentproject.com/blog/?p=200>

The causal theory of structured analytic techniques suggests that each structured analytic technique is designed to encourage analysts to engage in broadening checks, actions in the analytic process that "slow the production of [an] analytic product and make explicit the sacrifice of efficiency in pursuit of accuracy" (Zelik 2007 *et al.*). One broadening check is falsification which requires analysts to challenge status quo thinking by searching for rival hypotheses, thus increasing the scope of analysis, and rejecting as many of these as possible. For example, the Analysis of Competing Hypotheses (ACH) is based on the philosopher Karl Popper's argument that rigorously challenging and rejecting hypotheses is necessary for extending knowledge (Popper 1953). Popper (1972, p. 265) summed up this argument when he wrote "whenever a theory appears to you as the only possible one, take this as a sign that you have neither understood the theory nor the problem which it was intended to solve." Popper's point is as simple as it is profound: knowledge grows through testing multiple hypotheses and evidence rigorously rather than one.

Another proposition of the theory is that the more broadening checks in the analysis, the more that the triangulation of judgments can occur, thus producing a more accurate judgment. Underlying this proposition is the theory of triangulation (Campbell and Fiske 1959; Campbell et. al. 1966; Denzin 1978; Cook 1985). The basis of triangulation theory comes from geodetic survey methods which states that to find a position on a map, a surveyor should rely on multiple bearing points to get a closer approximation of where the target position falls (Dunn 2012, p.16). The wider the distance between bearing points the better, as each helps the surveyor find the target position. Divergent points let the

surveyor know how far away from the target position; convergent points tell him he is getting closer. In intelligence analysis, the points are not GPS coordinates, but pieces of evidence, hypotheses, and perspectives, which together increase the rigor of analysis. With each new piece of information the analyst takes in more information, increasing the relative completeness of what is known about the analytic problem. Increasing relative completeness is necessary for achieving greater accuracy, the ultimate target of triangulation. This theory is explained in more detail in [Chapter 2](#).

1.4 ADOPTING EVIDENCE-BASED INTELLIGENCE ANALYSIS TO ANSWER THE RESEARCH QUESTIONS

Evaluating the theory of structured analytic techniques in the context of the Intelligence Reform Act requires an approach that focuses on “what works,” that is, whether the effort to improve analysis by using structured analytic techniques has been effective. One such approach is evidence-based practice, which has been implemented in fields ranging from medicine (Sackett 2000), policy (Davies *et al.*, 2001), to policing (Sherman, 1998), to counterterrorism (Lum *et al.* 2006). The promise of evidence-based practice has led some intelligence practitioners to call for evidence-based intelligence analysis (Marrin 2012; Chauvin and Fischhoff 2010; Pool, *et al.* 2009), however, these researchers have yet to define it in the context of intelligence analysis. In this research, I define evidence-based analytic practice as the deliberate effort to develop a robust evidence-base linking the use

of specific analytic methodologies to quality intelligence analysis. Underlying this approach is the goal of building a robust body of evidence through carefully constructed evaluations of using analytic techniques.

Structured analytic techniques provide one remedy to improving intelligence analysis and foreign affairs, generally. However, to determine if the techniques can address this problem two issues must be addressed: whether the techniques can be implemented and if they are effective, and if so, under what circumstances. To answer these questions an evidence-based approach was taken using semi-structured interviews, a survey, a systematic review of research on the techniques, and an experimental simulation.

1.4.1 Research Question 1: The Implementation of Structured Analytic Techniques

The first research question addresses implementation: How often are structured analytic techniques used in the IC and what factors affect their use? There have been several attempts to implement structured analytic techniques, such as the Global Future Partnership which, brought experts outside of the IC to use the techniques.¹⁰ Efforts at the National Intelligence Council (NIC) have also introduced the Alternative Futures Analysis technique to produce the *Global Trends* series of reports. However, the Global

¹⁰ Interview with Warren Fishbein, June 11, 2014.

Future Partnership and NIC are an exception. Other research suggests that structured analytic techniques are rarely used on the job (Johnston 2005).

Several variables might explain why analysts could be reluctant to use the techniques. For example, Moore and Hoffman (n.d.) argue quality of training in the techniques—an organizational variable—is important. They assert that the current curriculum does not take into account analysts' individual learning styles and insufficient time is given to practice the techniques. Marrin (2007) argues that belief systems of analysts are an important variable. Anecdotal accounts from the IC paint a portrait of analysts suspicious of the value of the techniques, especially older analysts, which Moore and Hoffman (n.d.) attribute to the inability of structured analytic technique proponents to “make a convincing case that [analysts] ought to try something new” (Moore and Hoffman n.d., p. 2). However, there is no research on how pervasive this view is in the IC. At the same time the IC is undergoing a demographic shift, it is also facing a veritable ‘catch 22’ from tasking requirements: the amount of current reporting has increased (Johnston 2005, pp. 26-27; Dixon and McNamara, 2008) while analysts are under more pressure to use structured analytic techniques which are *perceived* to require more time (Heuer 1999, p. 86).

In [Chapter 4](#), an attempt is made to answer the implementation question by using key informant interviews and a survey of a US intelligence agency with 80 analysts—the largest ever attempted. The informant interviews were held with expert members of the IC willing to share their knowledge of the analytic reform movement and intelligence methodology. These informants were selected through a purposive and snowball

sampling strategy. While the key informant interviews focused on the overall implementation of structured analytic techniques in the IC, an in depth field study of INR was used to examine the variables hypothesized to affect the use of the techniques. The study included both a survey of 80 intelligence analysts and follow-up interviews with 15 analysts to probe the interpretation of the study variables. The results of the survey were analyzed with descriptive and inferential statistics and integrated with qualitative data from the interviews.

1.4.2 Research Question 2: The Effectiveness of Structured Analytic Techniques

The second research question addresses the effectiveness of the techniques: do structured analytic techniques improve the quality of intelligence analysis and, if so, under what circumstances? The U.S. Government's *Analytic Tradecraft Primer* (2009) lists 12 "core" structured analytic techniques (see table above), including scenarios (under the name alternative scenarios analysis), Team A/Team B, and red teaming. This list is significant not because it is comprehensive—Johnston (2005) identified more than 190 techniques used in the IC— but because it has been used to create the curriculum of the analytic reform movement's seminars and training materials. Therefore, the list is a representative sample of techniques promoted in the reform effort. However, it is unclear to what extent these core techniques might improve the quality of intelligence analysis except for some scattered anecdotal evidence (Marrin 2012).

To begin to answer the effectiveness research question, a systematic review was conducted of the evidence on structured analytic techniques. Defined, a systematic review is an “explicit method to identify, select, and critically appraise relevant research” (Cochrane Glossary 2014). Such reviews are used in evidence-based policy to sum up the best available evidence on a specific question to determine “what works” (Campbell Collaboration 2014). While the systematic review is similar to its close relative, the literature review, there is an important difference: transparency. Unlike a traditional literature review, the researcher conducting a systematic review makes his procedures of his review explicit, such as specifying the criteria for why a research study is (or not) included. This transparency should decrease the possibility of “cherry picking” evidence that fits the researcher’s preconceived notions of the data (Cooper 2009). For example, at the beginning of a systematic review, the researcher must set the inclusion criteria, the rules for using research reports.

Studies for the review were drawn from within the intelligence studies literature and outside in numerous disciplines, such as policy analysis and business. Evidence from intelligence studies comes from de-classified documents from the IC, such as a use of the Team A/Team B exercise during the Cold War (Mitchell 2006) and research conducted by students at intelligence training centers and universities. While these sources are useful, there is also a large body of evidence outside the intelligence literature. This evidence is available in the form of research studies published in peer reviewed academic journals. However, it is important to note that not all evidence is equal; an anecdotal

impressions is less credible evidence than a randomized control trial. Therefore, this research also takes into the quality of evidence by assessing each study's research design.

To supplement the systematic review, an experiment was conducted to evaluate two particular structured analytic techniques: ACH and Indicators or Signposts of Change (henceforth: Indicators). The Indicators technique requires analysts to list "observable events that one would expect to see if a postulated situation is developing" (U.S. Government 2009). In practice this might mean analysts listing the indicators of an upcoming coup, for example (e.g. the presence of rioting, political assassinations). ACH differs from Indicators in that it includes Karl Popper's idea that knowledge should advance through a process of conjectures and refutations, a process of "falsification." The use of ACH is fairly straightforward: analysts start by creating a matrix and then insert evidence in the rows and hypotheses in the columns. Next, each piece of evidence is compared with each hypothesis to attempt to determine what Heuer refers to as "diagnosticity"—the extent to which each of piece of evidence is consistent (or inconsistent) with each hypothesis. In generating evidence, analysts are encouraged to try to disconfirm the hypotheses. Once the matrix is complete, hypotheses that can stand up against the evidence remain. It is in this falsification process that ACH differs from Indicators. However, an open question is the extent to which ACH helps analysts falsify, rather than confirm their own beliefs, a phenomena found in both the psychological (Wasson 1968) and intelligence studies literatures (Tolcott *et al.* 1989).

The experiment was a forecasting simulation involving 21 foreign graduate intelligence studies students from the University of Pittsburgh randomized into two

experimental groups, ACH or Indicators group. Within each group were three analytic teams made up of 3-5 students. The task of the simulation was a realistic intelligence analysis task: study participants were asked to provide a percentage of chemical weapons that would be removed and destroyed from Syria as per the United Nations Security Council resolution for that country to destroy its stockpiles. Along with their predictions, participants also provided a narrative so that rigor could be assessed, and completed a short cognitive reasoning style questionnaire. To determine what extent structured analytic techniques improve over intuition, each group (the experts, students using ACH, and students using indicators) made intuitive judgments without using a structured analytic technique. However, after their initial intuitive judgment, student teams spent three hours analyzing the task using their assigned technique and made another set of analytic judgments.

Research from diverse literatures, such as social psychology (Thompson and Wilson 2014) and forecasting (Armstrong 2006) suggests that face-to-face collaboration can lead to social conformity which can in turn reduce the quality of analysis. Therefore, it was expected that the hypotheses participants generate in aggregate before collaborating face-to-face, will be greater than after the number of hypotheses generated after face-to-face collaboration, regardless of the technique used. This study hypothesis is particularly important for ACH because the technique's procedures call for analysts to collaborate and generate a full set of plausible hypotheses (Heuer and Pherson, 2011, p. 32).

This experiment also addresses how cognitive reasoning style might interact with the use of the techniques. Cognitive reasoning style is important in intelligence analysis because if an analyst has a more open style s/he is more willing to collect more disparate information and therefore likely to triangulate down to the correct answer in foreign affairs analysis (Tetlock 2005; Bar-Joseph and McDermott 2008). However, it is not clear to what extent, if at all cognitive reasoning style will interact with techniques on the rigor and accuracy of analysis.

1.5 AT THE FRONTIER OF FOREIGN AFFAIRS ANALYSIS

The analytic reform movement presents an opportunity to determine to what extent foreign affairs judgment can be improved. Tetlock's (2005) finding that expertise has clear limitations and a diminishing return on improving judgments of foreign affairs confirms what many in political psychology have long suspected: in the complex, chaotic international realm experts do little better than well-informed observers in making mid and long-term forecasts. Compounding the problem, foreign affairs is an area of policy making where being wrong is costly in both blood and treasure. Perhaps no other institution understands this fact as the world's largest producer of foreign affairs analysis, the United States intelligence community.

Fortunately, there are possibilities for improving foreign affairs analysis through structured analytic techniques. However, there are two important questions. First, while

the techniques hold promise for improving the rigor and accuracy of analysis, anecdotal accounts suggest that analysts are wary of using them. The literature provides several reasons ranging from increasing time pressure to the availability of training. Second, there are significant knowledge gaps in how the techniques should improve analysis.

The [next chapter](#) synthesizes and reframes the literature on intelligence to set a testable standard for quality analysis and a theory of how the techniques might improve analysis. With the inquiry framed and structured, [Chapter 3](#) outlines the multi-method research design to answer the research questions. The empirical chapters, [Chapters 4, 5](#) and [6](#), present results of this research. In the [final chapter](#), the results from the empirical chapters are synthesized to develop a set of evidence-based principles for intelligence analysis. Additionally, the final chapter sets forth steps to implement the evidence-based principles to improve intelligence analysis in the IC.

2.0 CHAPTER 2: ASSESSING THE QUALITY OF INTELLIGENCE ANALYSIS: A SYNTHESIS AND REFORMULATION

The purpose of the analytic reform movement is to improve the quality of intelligence analysis (Immerman 2011, p. 180; Fingar 2008), although defining “quality” is controversial and conceptually difficult (Marrin 2012). Section 2.1 examines the standard measure of intelligence and decision quality from the literature, the presence of cognitive biases. An examination of the cognitive biases literature suggests that while cognitive biases certainly exist in intelligence analysis—or any kind of information analysis for that matter—it is a narrow and unreliable measure. In place of cognitive biases, a two component measure of analytic quality is suggested. This framework for evaluating analytic quality takes into account the accuracy and whether analysis complies with a set of established rigor standards, including elements related to cognitive biases. Using accuracy and rigor as a standard, section 2.1 concludes with a discussion of how there is a diminishing returns on expertise.

In Section 2.2 intelligence analysis is reframed as an “inexact science” (Rescher and Helmer 1959) that can be improved through structured analytic techniques which leverage expertise and make the analytic process explicit. Next, the section covers the theoretical justification for the techniques from Heuer and Pherson (2014) that claims structured analytic techniques lead to system 2 (slow thinking) that debias system 1 (fast

thinking). However, this account misinterprets a key finding in the psychological literature: system 2 can be susceptible to the same unconscious screening and manipulation of information as system 1 (Kahan *et al.* 2006, Kahan 2013).

In section 2.3 a theory of how structured analytic techniques improve intelligence analysis is proposed based on a theory of triangulation (Campbell and Fiske 1959; Webb, Campbell, Schwartz and Sechrest 1966; Denzin 1978; Cook 1985). Specifically, I argue that the techniques require analysts to engage in a set of broadening checks that widen the scope of the analysis. As the analysis widens, analysts are able to triangulate to make more accurate judgments. It should be noted that while this theory is explanatory it is also normative in that it provides potential guidelines for how to improve intelligence analysis. In other words, it is a theory of a method.¹¹ The final section (2.3) of this chapter traces the intellectual history of structured analytic techniques and the eventual wide-scale implementation of these techniques as a key feature of the analytic reform movement. An important question about the success of the movement centers on the extent to which structured analytic techniques have been implemented. The literature on knowledge utilization and intelligence studies is synthesized to identify several variables – training, time pressure, and the demographic shift in the IC.

¹¹ A similar example would be game theory. In game theory certain assumptions are made and there is a desired goal or end state, to make or foresee another player's utility maximizing decision.

2.1 WHAT IS 'QUALITY' INTELLIGENCE ANALYSIS AND WHY IT'S SO HARD TO PRODUCE?

The mission of the intelligence community (IC) is to generate intelligence analysis to guide decision makers. However, a critical question is what constitutes analytic quality. The common determinant of quality in the international relations and intelligence studies literature is whether cognitive biases were mitigated or reduced but follow-up research on cognitive biases casts some doubt on this as a reliable measure of intelligence quality. Most notably, researchers have been unable to replicate many biases, even in lab settings (Hammond 1996, pp-203-213).

Instead, an alternative measure is a modified version of Hammond's (1996) two part measure that takes into account whether the analysis was sufficiently rigorous, such as the exploration of multiple hypotheses and verification of evidence, and also whether the analysis is empirically accurate.¹² Applying this measure as a benchmark of analytic quality, a probe of the literature suggests that expert judgment provides *diminishing* benefits for reliably generating quality intelligence analysis. In other words, expertise matters but not as much as conventional wisdom would lead us to believe. Further complicating matters, the use of heuristics in expert judgment, the limitations of an

¹² Assessment of empirical accuracy in foreign affairs tasks is undoubtedly difficult but possible. For a discussion of the challenges of evaluating empirical accuracy, see: Jay Ulfelder, "Jay Ulfelder on the Rigor-Relevance Tradeoff" The Good Judgment Project, May, 2014, available at: <https://goodjudgmentproject.com/blog/?p=200>

individual perspective (sometimes referred to as “mental models”), and the sheer complexity of international affairs appear to limit the effectiveness of expert judgment.

2.1.1 Defining Intelligence Analysis Quality

Each day the (IC) collects enough data to fill the Library of Congress—the largest repository of public knowledge in the U.S.—several times over (Aid 2013). This raw data is processed by approximately 20,000 government analysts plus a larger but unknown number of contractors funded by an estimated 75 billion dollar annual budget (Priest and Arkin, 2011). Central to this process is intelligence analysis, which one of its founders called the “[application] of the instruments of reason” to inform national security decision making (Kent 1949). To conduct analysis, intelligence analysts apply their expertise and reasoning to a myriad of data sources classified under a bewildering array of acronyms, such as HUMINT (human intelligence), SIGINT (signals intelligence), OSINT (open source intelligence), to name a few (Krizan, 1999). The products of this massive system are intelligence reports, disseminated to decision makers in a variety of mediums including presentations, briefing papers, and memos. Garst (1989, pp. 5-7) identifies six types of intelligence products including research, current, estimative, operational, scientific/technical, and warning. These products differ greatly in their purpose with some focusing on future or potential events (e.g. estimate and warning), while others can focus on short or near-term events (current).

However, the intelligence analysis process does not always produce “quality” intelligence analysis and in fact, even a cursory reading of the literature provides a long

list of cases from Pearl Harbor (Wohlstetter, 1962) to the Iraq War (Jervis 2006) where intelligence analysis errors – at least in part – led to foreign policy disasters. In analyzing these events, intelligence studies and foreign affairs scholars have retrospectively judged failure in terms of cognitive biases and errors of reasoning. For example, Jervis (2006) concluded the IC failed to question its assumptions about Iraq's weapons of mass destruction program and did not give due diligence, to invoke the famous Sherlock Holmes case, to the "dogs that did not bark." In other words, Jervis claims the IC fell prey to confirmation bias, the error of seeking to confirm rather than disconfirm one's beliefs.

Jervis' and other scholars work on cognitive biases are based on Nobel Prize winning research by cognitive psychologists Daniel Kahneman and Amos Tversky. Kahneman and Tversky used lab experiments to show that subjects do not think according to the tenets of rationality—extensive information search and proper weighting of utility functions. Instead they used mental shortcuts called "heuristics," and in the process, violated rules of probability theory or formal logic. These violations of probability theory and formal logic are called cognitive biases. In the 1970s, this research was imported to intelligence analysis (Heuer 1999) and international relations (Jervis 1976) as a basis for evaluating foreign affairs decision making and intelligence analysis. However, in the 1980s, long after Kahneman and Tversky's cognitive bias framework had migrated to the international relations and intelligence literatures, other cognitive psychologists began questioning the internal and external validity of their findings,

criticisms that are rarely discussed in international relations or intelligence literatures to this day.¹³

In the 1980s, after the cognitive biases literature had been imported into international relations and intelligence studies, psychologists found that if they slightly tweaked experimental conditions many of the cognitive biases would disappear. For example, Nisbett (1980) instructed test subjects in basic probability theory and found that so-called innate errors of human reasoning melted away. Also, other researchers have questioned whether the abstract reasoning tasks that Kahneman and Tversky employed to ‘expose’ cognitive biases are generalizable to applied settings such as high-level political decision making (Suedfeld and Tetlock 1992) and intelligence analysis (Moore and Hoffman n.d.). Tversky’s and Kahneman (1982) study of representative bias speaks to the abstract nature of many of the tasks—in the experiment they asked subjects to: “consider the letter R, Is R more likely to appear in the first position of a word or the third position of a word?” To date Kahneman has not addressed these criticisms. This follow-up research does imply that cognitive biases do not exist—it is undeniably true that humans are boundedly rational and make errors in judgment (Simon 1972). Rather, the problem is whether using cognitive biases as a psychometric measure based on probability theory and formal logic as a benchmark is a reliable measures of analytic performance.

¹³ For an extensive discussion of the methodological foundation of the cognitive biases research, see Hammond (1996) pp-203-213.

An alternative approach is to examine whether analytic judgments confirm to the established rules and norms of analysts and whether they are empirically accurate. Zelik, *et al.* (2010) explored what counts as the norm for quality intelligence analysis by surveying and observing analysts and developed a multi-attribute scale of depth covering eight attributes, which, when combined, reveal a composite assessment of analytic rigor derived from practitioners (see Table 2.1, below). For example, a key issue for attaining analytic depth is the exploration of alternative hypotheses, which subsumes elements of the cognitive bias, confirmation bias. Other depth attributes include the validation and extent to which diverse information sources were sought out, among five other depth attributes.

Table 2.1: Truncated and Reproduced Rigor Attributes from Zelik *et. al.* (2010)

| Rigor Attribute | Indicators of Rigor | |
|------------------------|---|---|
| | 'Shallow' | 'In Depth' |
| Hypothesis Exploration | Little or no consideration of alternatives | Significant generation and consideration of alternative explanations |
| Information Search | Failure to go beyond routine and available data sources | Collection from multiple data types |
| Information Validation | General acceptance of information at face value, no clear establishment of underlying veracity | Systematic and explicit processes employed to verify information |
| Stance Analysis | Little consideration of the views and motivations of source data authors | Perspectives and motivations of other authors/sources are considered |
| Information Synthesis | Little insight with regard to how analysis relates to the broader context or to more long-term concerns | Extracted and integrated information in terms of relationships rather than components and with a thorough |

| | | |
|--|--|--|
| | | consideration of diverse interpretations of relevant data. |
|--|--|--|

A more difficult issue is defining what counts as “sufficient” analytic rigor. Due to the sheer diversity of tasks analysts undertake, to recommend high rigor in all attributes across all is setting the bar unrealistically high (Zelik *et al.* 2007). Consider an analyst working alone on current reporting, an analytic task with short turnaround (less than a day): is it realistic for this analyst to achieve high depth across the eight attributes? Most likely not. To make sense of what counts as sufficient rigor, we can factor in the characteristics of the analytic task. Greitzer (2004) and Scholtz and Hewett (2004) surveyed analysts to provide analytic task characteristics, with two important for determining sufficient depth: complexity/breadth of topic and available time. Relying on these characteristics it is possible to determine a “rule of thumb” for sufficient rigor. For example, a complex task with a project length of several months, such as a long-term strategic forecast, would require rigorous analysis across the eight attributes to be considered sufficiently rigorous.

In addition to rigor standards, analytic quality can be evaluated by determining if an analyst ‘got it right.’ This measure is referred to as empirical accuracy because it is the extent to which an analyst’s judgments correspond with event in the observable world (Tetlock 2005; Hammond 1996). For example, if an analyst concluded the Soviet Union would fall in the late 20th century, and foresaw the Arab Spring he would score high on the correspondence measure. To date, the largest application of the empirical accuracy standard was Tetlock’s (2005) study in which he asked more than 200 foreign affairs

experts to make thousands of judgments over several years. In intelligence studies literature similar attempts have been made, albeit on a smaller scale. Feder (1995) reported a similar study when he disclosed the results of a forecasting model used at the CIA called Policon which was reportedly twice as accurate as a team of CIA analysts in forecasts of political instability. In addition to this work, other researchers have assessed correspondence by abstracting historical events to hide key details. For example, Folker (2000) provided abstracted vignettes of real events, such as the Allied invasion of Nazi Europe during World War II, to test how analysts utilizing an analytic methodology fared against those who did not.

2.1.2 The Diminishing Returns of Expert Judgment

The traditional approach to intelligence analysis, expert judgment, is based on the intuitive judgments of analysts, often working alone, in a variety of regional and functional areas (Johnston 2005). In its purest form, this mode of analysis relies on intuitive reasoning on the basis of expertise (Marrin 2012; Khalsa 2009). In addition, the traditional form of analysis is also understood to be a solitary activity, with most collaboration occurring after or at the end of the analytic process, rather than throughout.¹⁴ The solitary nature of expert judgment can be problematic because a single perspective is unlikely to subsume enough important information to conduct in depth

¹⁴ Personal communication with Richards Heuer Jr. (August, 2014)

analysis of complex issues (Churchman 1971). This threat is discussed in the intelligence literature through the concept of mental models, “the distillation of the intelligence analyst's cumulative factual and conceptual knowledge into a framework for making estimative judgments on a complex subject” (Davis 1992). Depending on the mental model of the analyst he or she will filter information differently, potentially excluding key information, such as hypotheses or pieces of information. Muller (2007) explores this concept through political affiliation and points out conservative and liberal analysts are likely to filter information differently because each has a different mental model, giving each side an incomplete picture of international threats.

Nowhere else has the limitations of expert political judgment been more apparent than in empirical accuracy. In one study conducted by the *Economist*, finance ministers, chairmen of multinational corporations, Oxford University economics students and a control group of London garbage collectors, were asked to make economic predictions from 1984 to 1994. While the number of respondents was low, just a dozen or so, the final results were shocking: the groups were very similar, and surprisingly, the garbage collectors were tied with the corporate chairmen. Tetlock's (2005) study replicated the *Economist's* experiment on a much larger scale with 200 experts and found that expertise matters, but only to a point. Interestingly, the experts easily beat out undergraduates but Tetlock found “highly educated and experienced experts...were not superior to untrained readers of newspapers in their ability to make accurate long-term forecasts of political events” (Kahneman and Klein 2009, p. 520). This result suggests there is a

diminishing returns on foreign affairs expertise; being an expert on a region or topic may improve accuracy over novices but not well-informed observers.

Two factors explain why expertise can only take us so far: the complexity of forecasting in international affairs and the lack of opportunities to learn (Kahneman and Klein 2009). Across disciplines, researchers have long understood that tasks where professionals make judgments involving human behavior, empirical accuracy is lower. Professionals in these areas include probation officers, counselors, intelligence analysts, and many more professions. Shanteau (1992, 1987) noted this distinction by differentiating between professionals who judge ‘things’ rather than people, such as the rotations of the moon or the change in the tides. Underlying the difference between these two types of professions are signposts in the environment that guide decision makers to form their judgments. These can take many forms ranging from cumulus clouds signaling an incoming thunderstorm or rioting as a sign of regime collapse. Common to all such signposts is that each gives decision makers a sense of what is or will happen in the judgment task: the warm front indicates to the weathercaster rain could be coming; reports of civil disorder indicate to the intelligence analyst regime change could be coming. However, not all signposts are created equal; some are more reliable and linear than others. This point that has been explored extensively by researchers using Brunswik’s (1952) lens model, a theoretical framework that models and measures how judges use indicators, informational cues in the environment. Karelaia and Hogarth (2008, pp. 415 and 420) summed up a sizable chunk of lens model research in a meta-analysis of 249 studies and found, not surprisingly, that accuracy is lower when

indicators are less reliable and nonlinear. This suggests that in certain intelligence tasks, such as determining the intentions of a terror group, where there are few reliable and linear cues, accuracy is likely to be low.

In addition to the casual complexities of intelligence analysis, there are also few opportunities for analysts to receive feedback. This point is important because decades of cognitive psychology research suggest that tasks with more opportunities for feedback will enable practitioners to be more accurate (Kahneman and Klein, 2009). For example, Chase and Simon (1973) studied chess players and found that chess masters leverage 10,000 hours of practice to develop increasingly more accurate judgments, a point popularized in Malcolm Gladwell's *Blink* (2005). The underlying factor driving the relationship between feedback and accuracy is that a practitioner can learn from their mistakes and re-calibrate for the next judgment.¹⁵ However, in intelligence analysis there are usually few opportunities to get feedback because as Tetlock and Mellers (2009) note, while "there are repeated opportunities for learning how to direct Predator-drone attacks" there are not many opportunities for "predicting the outcome of the quirky third-generation dynastic succession in North Korea." Therefore recommendations to give analysts feedback on their judgments might be feasible in some intelligence tasks but less so in others (Rieber 2004).

¹⁵ Beyond issues of intelligence and foreign affairs judgments, failure and feedback are instrumental in engineering and the design sciences. Petroski (2008) argues that in order to improve performance, designers should look to past failures rather than successes. For an extended discussion with examples, see: Petroski, Henry. *Success through failure: The paradox of design*. Princeton University Press, 2006.

Post-mortems of the most recent intelligence failures related to September 11th and the Iraq weapons of mass destruction assessments suggest that, among other factors, the traditional approach to analysis based on expert judgment might have contributed to these disasters (Finger 2008; Laipson 2005). This traditional approach is certainly not always the “wrong” way to conduct analysis (especially depending on the type of analytic task), but the literature suggests that while definitive conclusions are not possible, there is a need for new thinking about how to improve the quality of intelligence analysis.

2.2 REFRAMING INTELLIGENCE ANALYSIS AS AN “INEXACT SCIENCE”

One way to improve intelligence analysis is to re-conceptualize it as “inexact science” (Rescher and Helmer 1959) that uses special methodologies called structured analytic techniques. These methodologies could improve analysis by leveraging and structuring analysts’ expertise. However, the literature provides few clues as to overall causal mechanisms that might lead to an improvement in analytic quality from the techniques. One explanation is Heuer and Pherson’s (2014) claim that structured analytic techniques lead to system 2 (slow thinking) that debias system 1 (fast thinking). However, this argument is based on the incorrect assumption that biases cannot occur in system 2. In fact, research shows that the same kind of screening and distorting of evidence subconsciously in system 1 occurs in system 2 (Kahan *et al.* 2006, pp. 1093-1094). In other

words, both heuristic reasoning in system 1 and reflective reasoning in system 2 can contain biased thinking. An alternative theory is the theory of triangulation (Campbell and Fiske 1959; Webb, Campbell, Schwartz and Sechrest 1966; Denzin 1978; Cook 1985) that states that analysts using techniques will perform broadening checks to consider more perspectives, evidence, and hypotheses. Combining this information, analysts should be able to triangulate to make more accurate judgments.

2.2.1 Intelligence Analysis as an “Inexact Science”

Given the limitations of analysis relying mostly on expertise, intelligence analysis scholars and practitioners have explored the idea of shifting intelligence analysis from an expertise-based “art” to a structured, systematic “science” (Marrin 2012). However, many intelligence scholars and practitioners’ (Khalsa 2009; Kerbal 2008; Marrin 2007) reject a binary distinction between “art” and “science,” and instead conclude that intelligence analysis ought to be a mixture of both, as both systematic inquiry and expertise are necessary (Kerbel 2008). Unfortunately the intelligence literature provides little guidance on what the mix of “art” and “science” should look like in intelligence analysis. At this point, the literature seems to have hit a dead-end, unable to provide further explanation of how intelligence analysis is both art and science. Fortunately, the philosophy literature provides a useful conceptual framework.

An alternative framing of intelligence analysis is to define it as an “inexact science.” In their seminal article, “The Epistemology of the Inexact Sciences” Rescher and

Helmer (1959) argue there are two broad types of science, the ‘inexact’ type of applied disciplines and ‘exact’ type limited to highly abstract fields (Shanteau’s distinction between judging static “things” and people comes to mind). In the exact sciences the reasoning process is formalized takes place by formal logico-mathematical derivation of the hypothesis from the evidence.¹⁶ Consequently, Rescher and Helmer conclude there are few truly exact sciences, such as physics and mathematics.

The social sciences are inexact due to their reliance on a less formalized reasoning process and looseness of its predictions. Readers might notice that some areas of social science also include formalization. However, there is considerable debate as to whether highly formalized work (e.g. rational choice and game theory) has been used to the same effect in the social science as it has been in the exact sciences (Walt 1999). In other words, the inexact sciences can mimic the exact sciences by formalizing the reasoning process, but it is unclear if this improves predictive power. For example, a political scientist may use rational choice theory and statistical modeling to forecast an election, but his forecasts are likely more probabilistic than the astronomer estimating the orbit of a distant moon.

According to Helmer and Rescher, an added distinction of the inexact sciences is that expertise can play an important role. Such a pronouncement on the value of expertise might seem odd given Tetlock’s (2005) finding that there is diminishing returns in

¹⁶ Even in the cases of these so-called exact sciences, and as Rescher and Helmer note, there is likely a disconnect between how scientists do their work (called “logic-in-use”) versus how philosophers of science and others characterize the work of these scientists (called “reconstructed logic”). For an extended discussion of these logics see: Kaplan, Abraham. *The Conduct of Inquiry: Methodology for Behavioral Science*. Chandler Publishing Company, 1964.

expertise. However, Rescher and Helmer's argument still holds water because they argue that expertise, if structured and extracted carefully from experts, can make up for a major problem in the inexact sciences: the lack of stable and generalizable theory. According to Rescher and Helmer, theories in the inexact sciences are "quasi-laws" because they only apply to a particular time and place. For example, structural realism may explain well interstate conflict in the Western world from the 19th to the end of the 20th century well, but have less predictive power to understand conflict in prehistoric societies. Experts hold these quasi-laws as unarticulated background knowledge that can assist in specific contexts, but the problem is that this knowledge must be reliably extracted in a structured manner. For this reason, Helmer and Rescher conclude an important task for the inexact sciences is the development of special methods to leverage expertise, and potentially push the boundaries of the inexact sciences.

While never invoking Helmer and Rescher, a small group of intelligence analysis reformers have sought to implement special methods to structure analysis. Beginning in the early 1970s early intelligence reformers, such as Richards Heuer and Jack Davis began the hard work of introducing methodologies they termed "alternative analysis" designed to structure analysis and leverage expertise, just as Rescher and Helmer had suggested decades earlier. This new approach differed from the expert judgment approach in that it sought to make analysis more transparent, and hopefully, more accurate. To accomplish this task, alternative analysis drew techniques from other "inexact sciences", such as operations research and policy analysis. After the Intelligence Reform Act (2004) in which alternative analysis was specifically cited, intelligence instructor and researcher

Randy Pherson suggested the switch to “structured analytic techniques” as a way to get around the “alternative” terminology, which implied that analysts should use the techniques an alternative to traditional analysis.

More than 40 years after Davis and Heuer introduced the techniques to the IC, they have proliferated widely, if not thinly, with more than 150 techniques in use in the IC (Johnston 2005). However, in 2009 the IC published a primer of 12 “core” structured analytic techniques categorized into the three types of diagnostic, contrarian and imaginative.

Table 2.2: Structured Analytic Techniques

| | |
|-------------|--|
| Diagnostic | Key Assumptions Check Quality of Information Check Indicators of Signpost/Change Analysis of Competing Hypotheses |
| Contrarian | Devil’s Advocacy Team A/Team B High-Impact/Low-Probability Analysis “What If?” Analysis |
| Imaginative | Brainstorming Outside-In Thinking Red Team Analysis Alternative Futures Analysis |

The purpose of diagnostic techniques is to assess underlying assumptions, information, or hypotheses in an analytic argument. When performing diagnostic analysis, analysts should uncover weak components of their arguments. Diagnostic techniques were introduced as a way to move beyond analytic “fortune telling,” the

disparaging term for uncritical analysis based on unexamined assumptions and beliefs (MacEachin 1994). The techniques address assumptions by forcing analysts to consider the foundations of their arguments by writing and/or comparing pieces of evidence, assumptions, and hypotheses. For example, the Analysis of Competing Hypotheses (ACH), discussed in more detail below, assists analysts in disconfirming their favored hypothesis.

While diagnostic techniques assess the strength of arguments by examining the underlying structure, contrarian techniques force analysts to break arguments down by comparing competing arguments. For example, in Devil's Advocacy a designated individual argues against the conventional wisdom of a group. The potential benefits of contrarian techniques are twofold. First, challenging arguments and beliefs may encourage analysts to focus on weak areas to improve or clarify, similar to the diagnostic techniques. Second, contrarian techniques can be useful to "weed out" weak arguments.

In imaginative techniques, analysts are encouraged to consider new perspectives, futures, and ideas. Unlike diagnostic or contrarian techniques that attempt to converge on a more accurate answer, imaginative techniques generate several possible answers. In red team analysis, for example, analysts try to put themselves in the mind of an adversary. Considering how a dictator or terrorist plotter is thinking brings new perspectives to the table and can inform analysis. The purpose of imaginative techniques, such as Alternative Futures Analysis, is not to predict the future but provide the decision maker with multiple plausible scenarios, which can hopefully, lead decision makers to prepare for multiple futures. For example, the National Intelligence Council (NIC) use

the technique to produce the *Global Trends* reports which extrapolate key drivers, such as demographics, technology, and political issues, to guide strategic-level decision making.

2.2.2 A Justification for Structured Analytic Techniques: System 1 and 2 Thinking

While each of the techniques should improve the rigor and accuracy of analysis, the current justification for structured analytic techniques provided by Heuer and Pherson (2014) does not provide a coherent justification. Heuer and Pherson (2014) base their justification of structured analytic techniques on dual process theory from cognitive psychology. This theory suggests that cognition uses two systems: system 1 and system 2 (Kahneman 2011). System 1 covers the intuitive processing and as such it is extremely fast and unconscious, extracting information from easily accessible information stores containing knowledge. A common example of system 1 thinking is the intuitive reasoning used by a morning commuter; the commuter intuitively takes the same route to work each day. System 2 covers more slow and deliberative conscious thinking, such as that required to solve a difficult math problem. Heuer and Pherson argue that the problem of cognitive biases lies in system 1 because they argue that “system 1 is usually correct” but that all biases are the result of fast thinking (P. 4). Therefore, the answer lies in the usage of system 2. According to Heuer and Pherson: “structured analytic techniques are a type of system 2 reasoning designed to help identify and overcome the analytic biases inherent in system 1 thinking” (p.5). In short, using structured analytic techniques should

engender slow analytic thinking that should cause analysts to make fewer errors of reasoning or cognitive biases.

However, the distinction between system 1 and system 2 for debiasing might not be as clear as is commonly believed. Research suggests that the same kind of screening and distorting of evidence subconsciously in system 1 also occurs in system 2 (Kahan *et al.* 2006, pp. 1093-1094). In fact, a concept known as “defense motivation” can lead to subconscious screening of evidence and information that threatens one beliefs in both system 1 and system 2. As Kahan *et al.* state, “in effect, defense motivation biases individuals’ use of System II reasoning, causing them to use deliberate, calculating, and methodical analysis to support beliefs dominant within their group and to debunk challenges to those beliefs” (p. 1094). As a result analysts can use techniques to trigger reflective system 2 thinking, but there is reason to believe that this type of thinking will consistently debias.

A related problem for using the system 1 and 2 distinction is that structured analytic techniques use both systems, with some techniques relying more heavily on one than the other (Martin *et al.*, 2011). For example, some techniques rely on extremely formalized rule based processes that require system 2, such as ACH, among others. These are the techniques Heuer and Pherson seemed to have in mind when they make the claim about system 2 thinking debiasing. However, many structured analytic techniques “are not calculative in nature [and] instead rely on mental simulation [and] past experience” (Martin *et al.*, 2011, p. 33), thus relying on system 1. Examples of these techniques include many imaginative and contrarian techniques—all of which are in Heuer and Pherson’s

book (2010; 2014) — such as Red Teaming, Devil’s Advocacy, Team A/Team B, to name a few. Given that both systems are equally susceptible to biases and the techniques draw from both, the distinction between system 1 and 2 is not useful for explaining why the techniques improve analysis. To explain how intelligence analysis is improvable as an inexact science, a new theory is needed.

2.2.3 A New Theory of Structured Analytic Techniques

As discussed above, measuring the quality of analysis has two components: 1) it is sufficiently rigorous and 2) accurate. Analysis can be said to be rigorous when it is in-depth. Rigor is measured by attributes identified by Miller, *et al.* (2006) and Zelik, *et al.* (2007), such as the extent to which the analyst focused on a favored “pet hypothesis,” questioned the reliability of sources, among others. Analysis is also high quality when it is accurate, that is, when there is a high degree of correspondence between the analytic judgment and what happened in the external world (Tetlock, 2005, p. 10; Hammond 1996). Or in other words, the extent to which an analyst “gets it right.” An important prerequisite for effective use of structured analytic techniques is that analysts avoid premature cognitive closure. Cognitive closure is defined as the stopping rule at which an analyst will no longer consider new pieces of evidence, hypotheses, and so on (Wastell *et al.* 2014; Kruglanski 2004).¹⁷ If analysts reach closure early in the analytic process then

¹⁷ A major question is determining at what point an analyst can stop searching for new hypotheses and evidence. Dunn (1997) addressed this problem by defining a stop rule or limit to rival hypotheses as a

the techniques will not have the intended effect of broadening the analysis. Along with this important precondition this causal theory includes two propositions:

Proposition 1: *Structured analytic techniques increase the rigor of analysis through ‘broadening checks’*

Each structured analytic technique is designed to encourage analysts to engage in broadening checks, actions in the analytic process that "slow the production of [an] analytic product and make explicit the sacrifice of efficiency in pursuit of accuracy" (Zelik *et al.*). Such checks might include checking for multiple alternative hypotheses or validating information. Regardless of the check, each adds new information, increasing the scope and complexity of the analysis. The result of more broadening checks is potentially more rigorous analysis.

Each technique has strengths in different broaden checks so while some techniques might be more useful for considering novel perspectives, such as Red Teaming, others are more useful for testing hypotheses, such as ACH. According to Heuer and Pherson (2010, p. 34) "each technique may provide only one piece of a complex puzzle...multiple techniques can be used to check the accuracy and increase confidence in the analytic conclusion." Underlying this statement is the main assumption of multi-method inquiry: because each technique has its own strengths and weaknesses, analysts should select techniques with non-overlapping strengths (Diesing 1991, p. 90; Kaplan 1964). For

means to avoid premature closure as well as unlimited openness. For more information, see: Dunn, William N. "Pragmatic eliminative induction: proximal range and context validation in applied social experimentation." *PHILOSOPHICA-GENT*- (1997): 75-112.

example, an analyst might use an imaginative technique, such as Brainstorming to gain convergent hypotheses and a hypothesis testing technique, such as ACH to whittle down the possibilities.

Proposition 2: *The more broadening checks in the analysis, the more that the triangulation of judgments can occur, thus producing a more accurate judgment*

The theory of triangulation states that in inexact sciences like intelligence analysis, inferences will be more valid — that is, more accurate — if investigators triangulate among multiple data sources, perspectives, rival hypotheses, and hypotheses (Campbell and Fiske 1959; Webb, Campbell, Schwartz and Sechrest 1966; Denzin 1978; Cook 1985). In other words, the more broad and diverse the analysis, the more likely the inferences drawn from it will be accurate.

The first writings on triangulation come from Campbell and Fiske's (1959) discussion of measurement. Since there is not a 1:1 correspondence between a construct and measure, it is difficult to accurately measure with a single measurement. For example, measuring a complex construct like "terrorism" would require not only measuring the number of attacks and deaths, but also financial impact and psychological impact on the target audience. Underlying Campbell and Fiske's idea is the assumption of geodesic survey methods; if trying to find a particular point or position on a map, the surveyor relies on multiple bearing points to get a closer approximation of where the target point lies. (Dunn 2012, p. 16). In triangulation theory, the multiple points are not GPS coordinates, but bits of information; each new convergent or divergent piece of

information helps the analyst “zero-in” on the empirical reality. Two examples, multiple perspectives and hypotheses illustrate this point.

“One might recognize” write Heuer and Pherson (2010) that “accuracy is best achieved through collaboration among analysts who bring diverse viewpoints to the table” using the techniques (p. 6). Triangulation among multiple diverse perspectives echoes this statement as each perspective yields insights are not attainable with the others (Allison 1971; George 1975; Linstone 1989) and then combining each to determine where assumptions and arguments converge. Francis Galton provided the first empirical demonstration of triangulation of multiple perspectives while recording participants’ guesses of how much an ox weighed at a county fair. Galton aggregated the guesses and found that each successive estimate converged towards the true weight of the ox. Triangulation was driving this outcome: some guesses (“perspectives”) were higher, others lower, but the aggregate of guesses, the “wisdom of the crowd” converged on the true weight of the ox (Surowiecki 2004).

In addition to perspectives, structured analytic techniques allow analysts to test and triangulate multiple hypotheses. Chamberlin (1890, p. 756) noted that researchers must find a way to temper their “intellectual affections” by the testing of multiple rival hypotheses, plausible explanations for the data. While other researchers had addressed the need to test multiple rival hypotheses in the more exact sciences, such as geology (Chamberlin 1890; 1964), physics (Platt 1964), and the sciences, generally (Popper 1959), hypothesis testing in the applied, inexact sciences, such as intelligence analysis, is more troublesome due to little experimental control (Kerlinger and Lee 2000, p. 567). For

example, an investigator in a laboratory setting running mice through a maze, through careful research design, can immediately rule out many conditions that could explain the observed outcome, such as selecting (nearly) identical mice (e.g. age, sex, size, etc.). However, investigators in the inexact sciences lack the ability to control, a problem that Donald T. Campbell and Julian Stanley (1963) addressed using multiple rival hypotheses in another inexact science, policy analysis and evaluation. They argued that when studying complex phenomena where experimental control is low, investigators must take measures to account for external and internal validity. Within each type of validity are threats, which stand as potential rival hypotheses or explanations for the data. For example, an analyst might focus only a narrow set of cases of revolution that bias the analysis. This threat is called “selection” and occurs when systematic conditions in the case (or respondent) cause the observed effect.

This theory of structured analytic techniques based provides an alternative to Heuer and Pherson’s explanation. However, it also explains why the techniques might be difficult to implement: each additional technique, perspective, or hypothesis entails greater costs in time and resources (Cook 1985). In some areas of the IC, such as the National Intelligence Council, some techniques have been implemented. However, it is still unclear to what extent the techniques are being implemented and what variables affect their use.

2.3 IMPLEMENTING STRUCTURED ANALYTIC TECHNIQUES IN THE IC

Since the 1970s, the IC has sought to implement structured analytic techniques, although under different terminology (e.g. alternative analysis). Large-scale implementation has been unsuccessful but with each decade more emphasis has been placed on training analysts in the techniques. What is unclear is how certain factors might now be affecting the use of structured analytic techniques. Several potential barriers exist, such as the mixed quality of training analysts receive and perceived time investment of using the techniques. At the same time, another factor might be assisting in implementation: the demographic factor. As more Baby Boomers retire, more young analysts might be more willing to use the techniques.

2.3.1 The Answer is in the Head, not the Numbers

The implementation of structured analytic techniques can be traced to a CIA internal memo to explore how quantitative methods developed in academia during the 1960s could be applied to intelligence analysis. The result of this probe led to a volume, *Quantitative Approaches to Intelligence Analysis* (1978), edited by Richards Heuer. In the introduction Heuer (p.9) states that the quantitative methods in academia focus on “the kinds of problems that can be quantified” and therefore quantitative methods place a “rather severe and intractable limits on its applicability to the needs of government agencies concerned with foreign affairs, since most of the variables of interest simply

cannot be quantified.” O’Leary *et al.* (1974) conducted a similar probe to Heuer to determine the usefulness of quantitative methods for foreign affairs analysis. The authors were granted access to 545 documents from the State Department’s Bureau of Intelligence Research classified at “secret” or below and found after coding each of these documents for type of data used, such as categorical and continuous, the authors found that less than twenty percent contained quantitative data. As with Heuer, the investigators came to the same conclusion: quantitative methods, such as regression analysis, play a limited role in intelligence analysis. It was with this less than sanguine view that Heuer presented the results of the CIA study at the 1977 International Studies Association convention. After his presentation, Heuer was approached by Zvi Lanir, a senior officer in Israeli military intelligence. Similar to Heuer, he too was studying the role of quantitative methodologies for improving intelligence analysis, but Lanir was closely following the scholarship of Israeli-American scholars Daniel Kahneman and Amos Tversky. Lanir suggested “the answer is in the head, not the numbers,” that is, to improve analysis the IC should focus on understanding psychological factors (Heuer, 2009). After this chance encounter, Heuer began studying the then-blossoming fields of behavioral economics and cognitive psychology research and formed the cognitive biases framework.

A few years after his chance encounter with Lanir, Heuer developed ACH in the mid-1980s while teaching courses on counterintelligence. In the course Heuer asked his analysts if they had considered adversary deception and each time analysts would reply they saw no signs of deception. The problem, as Heuer pointed out to his analysts, was that they must focus on what is not readily apparent, because by definition deception

should not present tell-tale signs. (Heuer 2009). Given this problem, Heuer created ACH to encourage analysts to think about rival hypotheses, especially the deception hypothesis. Soon after teaching the technique in his counterintelligence courses, Heuer realized the technique was applicable to other forms of analysis. Today, ACH is probably the most utilized structured analytic technique and has seen many modifications (Pool 2010, p. 19). For example, Stech and Elsaesser (2005) automated ACH with Bayesian logic. Others have made broader refinements, such as Wheaton and Chido (2006) who created structured ACH, which allows analysts to simplify technique.

During the 1980s initial interest and use of structured analytic techniques remained low, although leadership by Robert Gates, the CIA Deputy Director of Intelligence from 1986-1989, brought greater attention to the analytic process, which in turn kept structured analytic techniques alive in the form of training materials. From his inaugural speech as DDI, Gates sought to make clear that he would not accept weak argumentation that did not differentiate between analysts' opinion and fact. In addition, Gates sought to bring in multiple perspectives to the intelligence process by bringing in outside academics and other non-CIA specialists (Davis 1999). However, Gates greatest impact on the analytic process was creating a stringent review process that increased the standard of evidence in analysis (Davis 1999)

2.3.2 New Era, New Analysis

The disintegration of the Soviet Union dramatically affected intelligence analysis, particularly the types of problems that analysts grappled with. Treverton (2009) argues the breakdown of the 2-bloc international system changed the composition of intelligence analysis problems from simpler “puzzles” to harder “mysteries.” While puzzles can be solved with more information, mysteries are difficult to solve even with information saturation. For example, a puzzle is how many nuclear weapons does North Korea have, but a mystery is what Kim Jung Un’s intentions are. As a result of a shift in problem types, the role of the quantitative methodologies Heuer examined in his 1978 lessened. One CIA analyst interviewed for this research who began his career in the 1980s corroborates this point: while he initially used mainly quantitative methods his work in the 1990s onward shifted to semi-quantitative and collaborative methods.¹⁸

As security threats changed from more puzzles to mysteries, the analytic process itself was transformed by shifting policy maker demands and advances in data collection. The speed at which analysis was conducted quickened during the 1990s as intelligence analysis went from more general analyses to being tailored for specific decision makers. For example, Davis (1997, p. v) notes the long analytic papers focused on the worldwide Soviet threat that were standard in the 1970s and 1980s gave way to a “combination of briefings and short but insightful written and multimedia products covering a broad

¹⁸ Interview with a CIA methodologist, June, 2014.

range of regional and transnational issues.” Second, new digital collection systems in the 1990s increased the volume of data so that the IC is now capable of collecting it on the scale described in the introduction of this chapter. The result is an IC under data “overload” at an ever increasing pace (Fishbein and Treverton 2004).

Against the background of changing international threats, a quicker intelligence cycle, and data overload, the IC suffered from several intelligence failures that engendered greater scrutiny of how the IC conducted analysis, and ultimately led to greater implementation of structured analytic techniques (or as it was known at the time, alternative analysis). One such failure was the IC’s inability to foresee the 1990 Iraqi invasion of Kuwait. According to Davis (2002) the central problem was analysts and policy makers held tightly to the assumption Iraq was still recovering from its bloody war with Iran, and that not challenging this belief led to shallow analysis, and ultimately, an inaccurate judgment. Concerned by this failure, Deputy Director of Intelligence Douglas MacEachin personally reviewed a large number of the CIA’s analytic products and found 1/3 had no discernable argument (Davis 1999, p. xviii). Echoing Heuer’s arguments in the 1970s and 1980s, MacEachin concluded the mind held unchallenged assumptions, that if not identified could lead to error.

To address the problem of implicit assumptions, MacEachin developed linchpin analysis, a diagnostic technique for checking assumptions. Since social science vocabulary was not favored among analysts, MacEachin gave alternative terms to the technique’s components: hypotheses were termed “linchpins” and assumptions “drivers.” The technique worked by analysts listing all underlying drivers and linchpins

on a piece of paper. Then the analyst could conduct a form of sensitivity analysis by examining how changing the assumptions might impact linchpins. If executed properly the benefits of linchpin analysis were a clarified chain of logic for the decision maker and examination of implicit assumptions. To implement linchpin analysis, MacEachin created “Tradecraft 2000,” a course that ran in the early 1990s with components that are still part of the Sherman Kent School’s curriculum, the CIA’s in-house intelligence training center that opened in 2000 (Marrin 2003).

In addition to the inaccurate estimates of Iraq’s willingness to invade Kuwait, the IC and how it conducted analysts faced withering criticism in the late 1990s. In 1998, the Indian government conducted a nuclear test, much to the surprise of the IC. At the time the strongly held assumption was that the current administration in New Delhi had little interest in following through with the test, despite campaign promises suggesting otherwise. In an IC review of the intelligence failure, the chair of the commission concluded that analysts failed to foresee the nuclear test because they assumed the ruling Indian party strategy was the same as Western party: make promises in the campaign and renege once in power (Select Committee on Intelligence 1998).

Little more than a year later, the IC was again faced criticism, this time accused of an inaccurate appraisal of the world-wide weapons of mass destruction threat. An outside commission run by Donald Rumsfeld concluded that analysts paid insufficient attention to denial and deception (Davis 2002). In particular, the commission argued that analysts should search for multiple hypotheses and seek to disprove the widely held hypothesis that challenger states such as North Korea were behind in their inter-

continental ballistic missile programs. However, Mitchell (2006) argues that the Rumsfeld commission, instead of representing a sincere attempt to critique IC analysis, reflects a wider effort by conservative hardliners to manipulate the intelligence analysis process to support the Republican position. Regardless of political motivations, the combined effect of the Rumsfeld commission with the Jeremiah report was an increase in the use of structured analytic techniques and the inclusion of more outside experts, (Davis 2002), a trend that would only accelerate after the September 11th attacks.

2.3.3 Intelligence Reform and Terrorism Prevention Act of 2004 and Large-Scale Implementation

While shifting requirements of the post-Cold War world and intelligence failures spurred greater interest in structured analytic techniques, the 1990s were but a prelude to what was to come. After the Iraq invasion turned up no weapons of mass destruction, the analytic practices of the IC came under tight scrutiny as panels, committees, and commissions initiated investigations and urged reform. The impetus for reform led to the Intelligence Reform and Terrorism Prevention Act of 2004 (henceforth: the Intelligence Reform Act). The sections relevant to structured analytic techniques in the Intelligence Reform Act (sections 1017 and 1019) set forth a mandate for the use of alternative analysis techniques --the earlier term for structured analytic techniques-- through the creation of an Integrity and Standards and Analytic Ombudsman positioned in the newly created Office of the Director of National Intelligence (ODNI). Richard

Immerman, a professor of history at Temple University was tapped for the position and tasked with ensuring that high standards of analysis are maintained throughout the IC, including the use of alternative analysis, as structured analytic techniques are termed in the legislation.

One of the first tasks for Immerman and his staff was creating common analytic standards across all of the IC's diverse 16 agencies through Intelligence Community Directives (ICDs), memoranda designed to provide guidance on policy and regulations to the IC. ICD 206 (2007) can be traced to the deficiencies in analysis found in the Iraq weapons of mass destruction estimate estimate, specifically the lack of sourcing. The directive requires that analysts consistently provide sourcing for their judgments. ICD 203 (2007) was formulated to "sharpen the critical thinking that underlay intelligence products" and included seven analytic standards, such as Objectivity, Independent of Political Considerations, Timeliness, Based on All Available Sources of Intelligence, and Exhibits Proper Standards of Analytic Tradecraft (Immerman 2011, p. 170). This directive provided the framework for a new class called "Analysis 101" for new intelligence analysts. At one level the purpose of the course is organizational, as it brings together analysts from all IC agencies to develop a common analytic vocabulary, but at a deeper level it is, of course, about improving how analysts draw inferences. For example, analysts "learn not to lock on to one hypothesis and then collect evidence that supports it but to brainstorm on the whole range of possibilities" (Kelly 2007). Towards the goal of improving analysis, Analysis 101 contains lessons on various types of structured analytic techniques (Immerman 2011).

2.3.4 Factors Affecting the Implementation of Structured Analytic Techniques

As of 2011 approximately 4,000 analysts, nearly a quarter of the IC workforce, have taken the Analysis 101 course (Defense Intelligence Agency 2011). Yet it is unclear if the coursework has led to greater use of structured analytic techniques. Additionally, it is not known what variables might affect the use of the techniques. These include organizational elements generally (Richard and Oh 1993) such as the existence of analytic training (Moore and Hoffman, n.d.) and the role of individual and organizational belief systems (Weiss and Bucuvalas 1980). For example, Marrin (2012) argues that in order for analysts to use the techniques they must be convinced they are effective for improving analysis.

Intelligence researchers and trainers, David Moore and Robert Hoffman (n.d.), argue the training has had little effect on how analysis is conducted because the curriculum does not take into account analysts' individual learning styles and insufficient time is given to practicing the techniques. For this reason Moore and Hoffman conclude that the IC's approach to instruction might reduce or "hinder the development of the skilled use (or even any use)" of structured analytic techniques. For Moore and Hoffman, the training regime itself must be overhauled if analysts are to use the techniques.

In addition to the issue of training, another factor affecting the implementation of structured analytic techniques is the role of time pressure. One anonymous analyst interviewed by Moore and Hoffman (n.d., p. 2) claimed "practitioners are likely to be

punished for critical thinking” and presumably using structured analytic techniques, “because it takes longer and looks and feels like academic dithering in an atmosphere heavily weighted toward fast-paced “production” using computerized tools rather than gray matter.” This point has been backed up by unclassified ethnographies in the IC. Both Dixon and McNamara (2008) and Johnston (2005) suggest analysts are reluctant to utilize techniques because of the increasing current reporting requirements that emerged in the 1990s and accelerated after September 11th (Johnston 2005, pp. 26-27). A survey by the Rand Corporation supports these findings: of three dozen analysts and managers, 30 percent identified time pressure as a threat to conducting quality analysis (Treverton and Gabbard 2008). At same time many techniques require a perceived, significant time investment to decompose a problem and externalize the information to paper or a screen (Heuer 1999, p. 86). However, Heuer and Pherson (2011)—both proponents of the techniques--argue that the time investment in techniques is well worth it, and in fact, can shorten the turnaround time for analysis over the long-term. For example, according to some FBI analysts, ACH saves time because it acts as a written record of important facts about a case that can be retrieved quickly by pulling up the matrix.¹⁹

At a more fundamental level, an implementation factor is the perceived value of the techniques. Anecdotal accounts from the IC paint a portrait of analysts wary of the value of the techniques, especially amongst older analysts, which Moore and Hoffman (n.d., p. 2) attribute to the inability of structured analytic technique proponents to “make

¹⁹ Private communication with Richards Heuer (August 2014)

a convincing case that they ought to try something new.” However, it is not known how widespread this view is held in the IC because there is no study or survey including a question directly addressing the perceived value of structured analytic techniques, although two studies present related information. Treverton and Gabbard’s (2008) survey of an unreported number of analysts and managers found that almost a quarter of respondents identified inadequate “tools” as a major problem, however, the term “tools” is ill-defined and may include information technology tools, such as databases. Marrin (2012) conducted an informal survey via the International Association for Intelligence Education’s list-serv of more than 30 intelligence studies scholars, including some former practitioners, and found that most respondents were wary of the value of structured analytic techniques. In short, there seems to be a suspicion that structured analytic techniques do not improve intelligence analysis, which Marrin (2007) argues “can change if the use of structured methods can be demonstrated to provide clear benefits to the practitioner...”

Despite the training difficulties, time pressure, and low perceived value of the techniques, the implementation effort might be gaining a boost from a greening IC. Since the early 2000s the IC has reflected the wider demographic trends of the US workforce with Baby Boomers retiring in ever large numbers. In addition, the growth of the IC after the September 11th attacks has increased the number of new hires, and presumably, younger analysts to fill entry-level positions. Whatever the cause, the numbers suggest an IC in transition: approximately 11,000 analysts of the total 20,000 were hired after September 11th. This trend is important because some observers, such as Immerman

(2011) argue 'Generation Y' is more "comfortable with, and open to, new techniques that enable collaboration and integration." Fingar (2011, p. 24) reflects this optimism by noting that persuading new analysts to "adopt new techniques and to work differently than the generations they are succeeding is easy." The intellectual flexibility and collaboration of young analysts is frequently juxtaposed against the more senior, "grey beards" in the upper management who received different training and were enculturated in an IC oriented towards Cold War threats.

2.4 SUMMARIZING THE ARGUMENT

Evaluating intelligence analysis requires a reliable and appropriate measure of analytic quality. In this chapter, a new two part measure of quality focusing on analytic rigor and accuracy was proposed. Applying this standard to intelligence analysis, the expert judgment model of intelligence is unlikely to reliably produce high quality analysis. Decades of research suggests that two factors, the limitations of analysts' mental machinery and the complexity of the international affairs, limit the role of expertise in foreign affairs intelligence analysis. Perhaps the best illustration of the diminishing returns of expertise in foreign affairs was demonstrated by Tetlock (2005) who found that the best experts are as likely to be accurate as the flip of a coin.

One approach to improve intelligence analysis is to re-frame it as an inexact science requiring structured analytic techniques to leverage and reliably extract the

expertise of analyst. While this reframing provides a general rationale, there is a lack of theoretical grounding for the techniques overall. Heuer and Pherson's (2014) theoretical justification, while notable as a first attempt to provide a justification for all structured analytic techniques, is flawed because it assumes that slow calculative thinking is superior to intuitive thinking. In fact, research suggests that both types of thinking are just as susceptible to screening and distorting evidence (Kahan et al. 2006, pp. 1093-1094). Additionally, Heuer and Pherson's rationale for the techniques overlooks the fact that the techniques use both types of thinking. To address the gap, a causal theory of the techniques was forwarded. According to this theory, structured analytic techniques deepen and expand intelligence analysis by getting analysts to consider more perspectives, evidence, and hypotheses through the use of techniques of varying strengths and weaknesses.

Synthesizing and reframing the literature on how the techniques might improve analysis is useful, but it does not address how they might be implemented in the IC. A brief review of the history of the techniques in the IC presents promise for implementation but also serious obstacles. On the positive side, there is increasing interest from reformers to implement the techniques on a wider scale with each successful high-profile intelligence failure over the last 25 years. The Indian nuclear tests and the Iraq War brought greater scrutiny on analytic practice and calls to formalize analysis. However, the watershed moment thus far is the Intelligence Reform Act, which mandated the training and use of the techniques. Still, these hopes have been tempered

with the everyday limitations of analysis in the IC, including time pressure and the reticence of some analysts to try a new approach, among other obstacles.

Turning the inquiry to applying the framework developed in this chapter to evaluate the implementation and effectiveness of the techniques is the subject of the next chapter. Doing so presents a formidable but not insurmountable challenge.

3.0 CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY

The eminent baseball statistician Bill James once wrote “Every form of strength is also a form of weakness.” (Lewis 2004, p. 257). In James’ quote are the core assumption of multi-method inquiry: each method’s strength is also its weakness. For example, statistical methods might increase generalizability of the inference through probability theory, but generalizability is purchased at the steep cost of depth of understanding. The opposite could be said of ethnographic methods; the depth of understanding provided by the method reduces its generalizability.²⁰ This research project incorporated the logic of multi-method inquiry by leveraging the strength of numerous methods to build “interlocking patterns of converging evidence” (Tetlock (2005, p. 7).

This chapter lays out a blueprint for addressing the implementation and effectiveness of the techniques (see Table 3.1). Several variables have been identified that might impact the use of the techniques including the perception of the techniques (Marrin 2007; Heuer and Pherson 2011, p. 337), and demographic characteristics of analysts, such as age (Immerman 2011; Fingar 2008), and training (Moore and Hoffman, n.d.). To address this question, semi-structured interviews were conducted with intelligence experts, and a survey and follow-up interviews were conducted at an intelligence agency (Section 3.1). The second research question addresses the effectiveness of structured

²⁰ For an explanation of the methodological costs of various methods, see William N. Dunn, *Public Policy Analysis: An Introduction*. Upper Saddle River, NJ: Pearson, 2012, pp. 18-19.

analytic techniques. To answer this question, a systematic review and field experiment were conducted (Section 3.2).

Table 3.1: Research Questions

| Issue | Question |
|----------------|--|
| Implementation | How often are structured analytic techniques used in the intelligence community and what variables affect their use? |
| Effectiveness | Do structured analytic techniques improve the quality of intelligence analysis and, if so, under what circumstances? |

3.1 FIELD STUDY: IMPLEMENTING STRUCTURED ANALYTIC TECHNIQUES

To answer the implementation question, a field study design was conducted, a nonexperimental inquiry to “systematically pursue relations and test hypotheses” without manipulating the study variables (Kerlinger and Lee 2000, p. 585). Previous field research in the IC has brought insights on a variety of topics such as analytic culture (Johnston 2005), the use of intelligence sharing tools (Dixon and McNamara 2008), analytic tradecraft (Treverton and Gabbard 2008), and civilian intelligence education (Spracher 2009). Building on this tradition, in the field study three methods were used: semi-structured interviews with intelligence experts, a survey of an IC agency, the State Department’s Bureau of Intelligence and Research (INR), and follow-up interviews with INR analysts. Each method was selected on complementary strengths. For example, since the strength of semi-structured interviews is in gaining in depth understanding, this

method was supplemented with a survey to gain greater measurement precision (Dunn 2008, pp. 18-19).

3.1.1 Semi-structured Interviews with Intelligence Analysis Experts

The focus of the semi-structured interviews were “key informants,” reflective members of the IC willing to share their knowledge of structured analytic techniques and the analytic reform movement (Bernard 2006, Campbell 1955). In this study, informants are termed “intelligence analysis experts,” defined by their experiences and knowledge. For example, informants had to have significant time in the IC (10+ years) during the years of the analytic reform movement (2004 onward).

Informants were selected through purposive and snowball sampling strategies. These strategies were used because of the focus of the inquiry on accessing expertise in the population rather than generalizability (Bernard 2006). The first step in the sampling strategy was to generate a list of approximately 10 intelligence experts in consultation with the assistance of Dr. Phil Williams, one of the committee members. Next, each expert was contacted to set up interviews either in person or over the telephone. Since most respondents were in the Washington, D.C. area, two trips were made there in February and June of 2014. Upon interviewing informants, each was asked to identify other intelligence experts so as to ‘snowball’ the sample, adding more informants. After both purposive and snowball sampling strategies, a total of five intelligence experts were identified. The remaining five did not respond to inquiries for interviews.

The main research instrument was a semi-structured interview protocol (see Table 3.1). This instrument was used to focus the conversation on variables related to implementation of the techniques while allowing for the freedom to examine emerging concepts (Lee 1999, p. 62). For example, one question asked “How often do you believe analysts are utilizing structured analytic techniques on the job?” Another question addressed how training impacts implementation of the techniques (“How well prepared do you believe analysts are for using structured analytic techniques?”). The approximate length of the interviews was 30-60 minutes during which detailed field notes were taken. Due to the sensitivity of the topic, I did not record the interviews.

Table 3.2: Selected Interview Questions and Variables

| Variable | Question |
|-----------------|--|
| Training | How well prepared do you believe analysts are for using structured analytic techniques? |
| Perceived Value | In your opinion, what is the perceived value of structured analytic techniques among the “average” analyst? |
| Time Pressure | What effect do you think increasing time pressure has on the use of structured analytic techniques? |
| Demographics | The intelligence community appears to be seeing a demographic shift, with an influx of younger analysts. Do you think this will have an impact on the use of structured analytic techniques? Why or why not? |

3.1.2 Survey and Interviews at the Bureau of Intelligence and Research

The semi-structured interviews with intelligence experts provide a good overall snapshot of the implementation of the techniques and the analytic reform movement, but the comparative advantage of interviewing over other methods is in depth of understanding,

not measurement precision. This weakness is especially pronounced in understanding the extent to which analysts use structured analytic techniques on the job as key informants' impressions can be highly fallible. To compensate for this weakness, a survey was conducted. However, conducting a survey in the IC is no easy task, as evidenced by the fact that, there is only one survey in the unclassified domain. Spracher (2009) conducted a survey of 30 new analysts at the Defense Intelligence Agency. The difficulty of conducting a survey in the IC stems from the fact that intelligence-related research problems are examples of "difficult to access" problems (Maravic 2012). These types of problems involve actors who seek to "guard their secrecy, conceal their activities, decide who is allowed (not) to know, and have no interest in being observed or understood by others" (Maravic 2012, p. 153). Common examples of difficult to access problems include corruption, terrorism, international crime, and, of course, intelligence agencies. Consequently, a gatekeeper was needed, an individual who could "vouch" for this research and provide access to areas and people otherwise denied to outsiders (Kenney, 2013, pp. 29-30). For this research I was fortunate—and grateful—to gain access to INR through another committee member, Dr. Shawn Bird.

The 80 respondents included in the final survey were intelligence analysts currently employed at INR or other employees who were formerly analysts, such as intelligence managers. For this research, an analyst is defined as an INR employee who, in his or her current or previous work, applies or applied their expertise to weight data and test hypotheses about important international events (Johnston, 2005, p. 3). To construct the survey, the tailored method was used. This approach uses "multiple

motivational features in compatible and mutually supportive ways to encourage high quantity and quality response to the surveyors request” (Dillman *et al.* 2011, p. 16). Underlying the tailored design approach is the belief that both researcher and respondent are engaged in an exchange relationship where each party seeks a benefit (Blau 1964). Flowing from this exchange relationship is the implication that the key to getting high quantity and quality responses lies in constructing surveys that increase benefits, while decreasing the costs of participation. For example, to increase the benefits of taking the survey, the invitation to participate for this survey, as well as the opening script, offered the respondent the opportunity to get the final results of the survey. To keep the costs of participation low, the survey was designed to be completed quickly, with the average response time for the survey ranging from 2-3 minutes.

The survey instrument contains 10 questions with 6 focusing on the main variables (see Table 3.2 below). These questions and the format of the questionnaire went through several revisions, with close consultation of the dissertation committee to increase validity and clarity. Another important issue is reliability, the extent to which different respondents understood question items similarly. Reliability of the survey was calculated using a test-retest reliability by having six respondents retake the survey two months later. Reliability of an instrument is high if there is a strong correlation between the question items between the first and second occasion the respondent took the survey. A strong correlation is 1 while no correlation is 0. The correlation between the questions regarding the use of the techniques was low-moderately strong (.42). While the survey question provided examples of structured analytic techniques, analysts might have been

unsure what counted as using a technique. Fortunately, other questions, such as whether the analyst received training, were perfectly correlated (1.0). In short, the survey instrument is moderate to highly reliable.

Table 3.3: Survey Variables and Questions

| Variable | Question | Scale | Reliability (0-1) |
|---|--|---|-------------------|
| Use of Techniques (Dependent Variable) | "How often do you use structured analytic techniques on the job?" | -Occasionally/ Rarely -Never | .42 |
| Preparedness/Training | "Several kinds of 'structured analytic techniques' are used by analysts. These include Analysis of Competing Hypotheses (ACH), link analysis, and red teaming, among others. Have you received any training in any of these structured analytic techniques?" | -Yes -No | 1 |
| Perceived Value of Techniques: Rigor | "To what extent, on average, do you think structured analytic techniques help analysts think in a more effective way (e.g. consider new perspectives, challenge mental models, etc.)?" | -A great deal -A fair amount -A little -Not at all | 1 |
| Perceived Value of Techniques: Accuracy | To what extent, on average, do you think structured analytic techniques help analysts be more accurate or "right" in their analytic judgments? | -A great deal -A fair amount -A little -Not at all | .86 |
| Time Pressure | "On average, how long does it take you to complete an analytic project?" | -A day or two -A week or two -A month or two -Three months or more | 1 |

| | | | |
|------------|--|---------------------|-----|
| Age Cohort | "How many years have you been an analyst?" | N/A (continuous) | N/A |
|------------|--|---------------------|-----|

The survey was distributed over email to provide a useful and cheap medium for accessing respondents (Dillman *et al.* 2011, p. 44). Respondents were identified via an internal email list of 350 INR employees which included analysts and non-analysts (to screen out individuals not meeting the respondent criteria, a question was included that asked the respondent for his or her job title). While the exact number of analysts at INR is not known, there are approximately 200.²¹ Implementation of the survey was done through two rounds, with the first running July 29-August 5, 2014 and the second running August 8- August 15, 2014. Over the two rounds, 137 surveys were initiated, including 95 with at least 80 percent or more of the questions completed. Of the 95 responses, 80 respondents could be described as analysts under their current or former job titles. The remainder included other non-analyst job titles, such as "office director" and "foreign affairs officer."

An important consideration is whether the sample can be generalized to the remainder of INR analysts. One potential obstacle to generalizing to the rest of INR is if the analysts that did not respond were different in meaningful ways from those that did respond; this is called nonresponse error (Dillman *et al.* 2011, p. 17). Perhaps the most important way that analysts responding could differ from those that did not, is in their use of the techniques. For example, only those who use the techniques might have been

²¹ Personal communication with INR analyst (January 23, 2015).

interested in responding to the survey. However, the results of the suggested this type of nonresponse error was avoided because those who reported not using the techniques are well-represented as approximately 30 percent of respondents reported never using the techniques.

Another issue is whether the sample size of 80 is sufficient to generalize to the remaining 200 analysts. To determine this number requires calculating the appropriate sample size for a population of 200 (see equation 1). The appropriate sample size (N_s) based on a set of parameters, such as the acceptable level of sampling error (B), the amount of confidence desired in estimates (C), the level of variation in the main characteristic (p), and size of the population from which the sample is drawn (N_p).

Figure 3.1: Formula for Estimating Sample Size (Dillman *et al.* 2011)

$$N_s = \frac{(N_p) (p) (1-p)}{(N_p - 1) (B/C)^2 + (p) (1 - p)}$$

For this research, the acceptable level of sampling error was set at ± 10 percentage points and the confidence was held at the 90 percent confidence level. Prior to implementation of the survey, the expected variation on the main variable of interest, the use of the structured analytic techniques, would be a roughly 50/50 split between users and non-users. Compiling this information into the formula (see equation 2) provides the appropriate sample size of 65, which is well below the 80 that completed survey, therefore allowing the use of statistical procedures to generalize to INR's entire analytic cadre, the first time a survey has been generalizable to an IC agency.

Figure 3.2: Calculation of Appropriate Sample Size

$$65 = \frac{(200) (.5) (1-.5)}{(200-1) (.10 / 1.65)^2 + (.5) (.5)}$$

The survey data was analyzed using chi square tests to determine if the study variables were related. The test is useful in two situations: when the data is categorical and the researcher is interested in testing the hypothesized relation between two categorical variables. Given the focus of the first research question on determining the link between categorical variables hypothesized to affect the use structured analytic techniques, the chi square test was the most appropriate for analysis of the survey data. In addition, a strength of this statistical test is that it does not assume normality; however, it does require that cell counts are not less than five. In order to keep cells above the five threshold, each of the chi square tests necessitated combined categories. For example, there was a low number of analysts who served during the Cold War era and interim period between the end of the Cold War and September 11th. Therefore these categories were collapsed into Pre-September 11th.

Survey respondents were asked if they would be willing to be interviewed for a follow-up interview. Of the 20 analysts that indicated that they would be willing to participate in a follow-up interview, 9 agreed to a 15 minute phone follow-up interview. Respondents were asked open-ended questions similar to those in the survey to gain more description of the study variables (see table 3.3). During the interview, detailed

notes were taken and uploaded into *NVivo* then analyzed using a grounded theory approach to understand the underlying processes affecting analysts' use of the techniques (Glaser 1978). Grounded theory is helpful when there is inadequate or little theory to explain a particular phenomenon (Creswell 2013). This is the case with theory on the implementation of structured analytic techniques; beyond some scattered literature and anecdotal accounts there is little available information. The data from the interviews, contained in a set of digitized field notes, were uploaded to the qualitative data analysis program *NVivo* (QSR International 2013). Next the data was analyzed through the method of 'constant comparison' (Glaser and Strauss 1967; Lincoln and Guba 1985). To use constant comparison researchers engage in a sorting process identifying specific ideas called "codes" and then through an iterative process aggregate these codes to more general "themes." For example, one informant stated that "you will find a lot of haters in the older generations" concerning the use of structured analytic techniques. This snippet of text was coded under the broader code, "A graying IC" and aggregated under the theme "individual-level factors." This process was repeated to generate a series of themes describing the interviews (see Methodological Appendix A for the complete coding tree).

Table 3.4 Follow-up Interviews Questions

| Variable | Question |
|--------------|--|
| Age Cohort | The intelligence community appears to be seeing a demographic shift, with an influx of younger analysts. Do you think this will have an impact on the use of structured analytic techniques? Why or why not? |
| Preparedness | How well prepared do you believe analysts are for using structured |

| | |
|--------------------------------------|--|
| | analytic techniques? Have you received training in structured analytic techniques? |
| Perceived Value of Techniques: Rigor | In your opinion, what is the perceived value of structured analytic techniques among the “average” analyst at INR? |
| Time Pressure | What effect do you think increasing time pressure has on the use of structured analytic techniques? |

3.2 EVALUATING STRUCTURED ANALYTIC TECHNIQUES: A SYSTEMATIC REVIEW AND FIELD EXPERIMENT

Evaluating the techniques in the IC is difficult due to the secrecy surrounding intelligence analysis. Therefore, addressing the question of if structured analytic techniques improve the quality of analysis required a careful selection of multiple methods and research designs. The assumption of the design is that if multiple sources of evidence from data sources converge that are similar to how intelligence analysis is practiced in the IC, then a strong inference can be drawn. A systematic review and an experiment were chosen to complement the strengths of the other. The systematic review synthesized the evidence on all the techniques but this wider perspective is purchased at the cost of specificity as looking at all the techniques in aggregate makes detailed observation of each difficult. To address this weakness, the experiment evaluates two specific techniques – Analysis of Competing Hypotheses (ACH) and the Indicators of Change or Signposts technique.

3.2.1 A Systematic Review of Structured Analytic Techniques

To generate a consolidated body of evidence on structured analytic techniques, a systematic review was conducted. Defined, a systematic review is an “explicit method to identify, select, and critically appraise relevant research” to determine “what works” (Cochrane Collaboration 2014). The power of compiling and synthesizing research was first grasped in medicine (Cochrane Collaboration 2014), and has extended to other areas, such as public policy, most notably in the form of the Campbell Collaboration. For example, a Campbell Collaboration systematic review by Lum *et al.* (2006) on counterterrorism policy research from 1971-2002, found that of the approximately 14,000 reports only 7 had moderately rigorous research designs.

What differentiates the systematic review from its close relative, the literature review? The answer is in the transparency of the method, beginning with how studies are selected and concluding with the synthesis of findings. In a literature review, the selection of studies is idiosyncratic and often unclear to the reader, which is problematic because the researcher may inadvertently—or even advertently—omit studies, and consequently, bias their results. In addition, the lack of clear procedures also makes reproducing the results of a literature review difficult, if not impossible. Yet, the problems with literature reviews do not stop there: they also lack the ability to determine what level of effect (if any) one variable had on another (Cooper 2010, p. 7) while taking into account the quality of the study design and therefore the strength of the inference that can be drawn about what was found.

Still, there are also considerable weaknesses that limit the applicability of systematic reviews as they are traditionally conducted, most notably the reliance on quantitative data. This limitation is especially problematic in fields where quantitative data is hard to come by as is the case in intelligence studies (Zegart 2007). To address this limitation, an alternative form of systematic review was used in this research, the case survey method (Lucas 1974; Dunn and Swierczek 1977; Pawson *et al.* 2005). While known by numerous names, this approach has been used to study a variety of different phenomena, ranging from urban policy (Yin and Heald 1975) to organizational change (Dunn and Swierczek 1977). Instead of a traditional systematic review that extracts only numerical data, the case study method extracts narrative descriptions of outcomes and converts them into numerical data using a coding instrument. In this way, the case study method overcomes the reliance on quantitative data.

The first step in a systematic review is the specification of the population of research studies, the unit of analysis. In this research, studies included both intelligence and non-intelligence analysis studies. The decision to include both of these study types was based on the assumption that intelligence analysis does not differ significantly from other forms of information analysis, such as business and policy analysis. While analysis in other fields is not identical to intelligence analysis, owing to the secrecy of sources and methods (Warner 2002), the general process is similar. For example in intelligence analysis, as with other forms of analysis, it occurs in complex, high stakes environments often under time pressure (Johnston 2003, pp. 62-63). For example, financial analysts might have to forecast highly uncertain outcomes—perhaps a high-impact low-

probability event, such as a market crash--where their analyses could sway a multi-million dollar decision. This argument and example suggests non-intelligence studies are *sufficiently similar* to be generalized and useful to intelligence practitioners.

The data source for the systematic reviews was Google Scholar because it has extensive coverage of peer-reviewed research. According to one empirical study, "Google Scholar covers 98 to 100 percent of scholarly journals from both publicly accessible Web contents and from subscription-based databases" (Chen 2010). Gehanno *et al.* (2013) corroborated this claim and found that Google Scholar covered 100 percent of a sample of medical studies. In addition to published research, Google Scholar also covers non-published, or "gray" literature in the form of technical reports and dissertations (Google Scholar 2014).

Search terms were generated using the names of the 12 structured analytic techniques from the *Tradecraft Primer* (2009). Other search procedures included: searching for keywords in the title only to focus on the most relevant studies, excluding non-English studies, patents, and citations listings. Six of the techniques did not return any results: Key Assumptions Check, Quality of Information Check, High-Impact/Low-Probability, Indicators of Signpost/Change, "What If?" Analysis, and Outside-In Thinking. The other techniques varied greatly in the number of search hits: Alternative Futures Analysis (753 hits), ACH (20), Red Teaming (21), Team A/Team B (17), Brainstorming (838), and Devil's Advocacy (31).

Since it was logistically difficult to examine all studies from Alternative Futures Analysis and Brainstorming, a random stratified sample was conducted, a sampling

strategy that involves the division of a population into smaller groups (known as strata) to ensure equal representation of an important attribute (Kerlinger and Lee 2000, p. 179-180). In this study, an important attribute was “study relevance,” and therefore strata were divided into high ranking and low ranking studies in the search results. To determine high-low ranking and define the strata, Beel and Gipp’s (2009) study was used. Beel and Gipp (2009) reversed engineered the Google Scholar search algorithm (it is a trade secret) and found that the search results conform to the Matthew Effect: 20 percent of articles consistently appear at the top of the search results. Therefore, to sample high and low ranked articles equally, the population was divided, into a high (top 20%) and low strata (bottom 80%). The sample size was determined by setting the confidence level at 95 percent confidence interval and a margin of error at ± 10 percentage points. For example, Brainstorming returned 838 hits and the sample size calculation at 95 percent confidence interval and a margin of error at ± 10 percentage points was 87 studies. Broken into the two stratum, there were 17 high ranked articles (the top 20%) and 70 low ranked articles (the bottom 80%) downloaded.²² This process was repeated for Alternative Futures Analysis and for the other techniques all studies were downloaded (and did not require sampling). The sampling and downloading process yielded 259 studies (out of a total 1,724 search hits) and required approximately 10 hours.

²² To randomly sample, the following procedure was followed: Using a random number generator to select two sets of numbers page number and study position. For example, if there are 20 result pages possible and there are 20 studies per page, the randomly generated number 0220 would require downloading the study from the 20th position on page 2.

Next, the intelligence and non-intelligence research reports were downloaded and abstracts read to determine if the study provided some report, verbal or numerical, of the technique's effect on rigor or accuracy. Studies also had to make an attempt to evaluate a specific, identifiable instance of a technique. As a result, hypothetical reports (e.g. "ACH could be helpful to intelligence analysts...") were excluded. Studies were also excluded for several other reasons. For example, 95 studies described how to use a technique in a particular application but provided no explicit evaluation and 36 studies discussed conceptual issues related to using a technique (see Table 3.4). The exclusion process yielded 46 evaluative studies and required approximately 10-12 hours to carefully read through the abstracts.

Table 3.5: Excluded Studies

- Was a book – only articles and short manuscripts were considered for logistical reasons (2 studies)
- Describes how to use the technique in an application but provides no evaluation of effectiveness (95 studies)
- Only discusses issues related to the use of the technique (36 studies)
- Duplicates a previous observation (e.g. multiple studies covered the IC Team A/Team B exercise from the 1970s).²³ (4 studies)
- Excluded studies looking at educational outcomes, such as creative writing skills.²⁴ (6 studies)
- Proposes an improvement or modification of the technique-but no evaluation (13 studies)
- In another language (not available in English) (1 study)

²³ In these cases, the study with the stronger design (operationalized by the study's MSMS score) was selected, if designs were identical the first sampled study was included, if there was still a tie, then the newest study was included

²⁴ While these outcomes are of partial interest, they are not closely related to outcomes related to intelligence analysis


- Was the comparison of one variation of a technique against another (28 studies compared one variation of a technique against another)
- Technique addressed in study was not one of those included in this study as defined in the *Tradecraft Primer* (28 studies)

In this systematic review, the relevant information was extracted from the studies and transformed into numerical values for analysis. For example, relevant information included, the research design of the study and the reported effectiveness of the technique on rigor and accuracy. Effectiveness was determined by coding each study according to the reported effectiveness of the technique on four levels: 1) the technique had a negative effect; 2) no effect; 3) mixed effect; and 4) positive effect.

To assess credibility, the internal validity of the 46 studies were assessed with the Maryland Scientific Methods Scale (MSMS) (Sherman 1998) which was modified to add a sixth level to include studies using meta-analysis (see figure 3.5 below). The least credible studies fall into MSMS levels 1 and 2. These studies either are a correlation between the use of a technique and a reported effect on rigor and accuracy (level 1) or are a simple pre-test and post-test design (level 2). More than half of all the evaluative studies, 24 in total, had low credibility. Moderately credible studies fall into levels 3 and 4. These studies build on level 2 research by adding a comparison group (level 3) and controlling for variables that might explain the outcome in the study (level 4); 7 studies were coded as moderately credible. Highly credible studies are those that either use random assignment or were systematic reviews employing meta-analysis, a statistical

technique used in traditional systematic reviews to synthesize the impact of the intervention. 15 studies were coded as highly credible.

Figure 3.3: The Maryland Scientific Methods Scale



| Level | Description | Example |
|-------|---|---|
| 1 | Correlation between a structured analytic technique and a measure of either rigor or accuracy measures | Analysts report using brainstorming generates more ideas |
| 2 | Measures of efficacy before and after using a technique, with no comparable control condition | Analysts generate more ideas after using brainstorming |
| 3 | Measures of efficacy before and after using a technique in experimental and comparable control conditions | Analysts generated more ideas in a treatment group but there was no improvement in efficacy in a control group |
| 4 | Measures of efficacy before and after the program in multiple experimental and control units, controlling for other variables that influence analytic performance | Analysts generated more ideas in a treatment group but there was no improvement in efficacy in a control group controlling for level of expertise and cognitive reasoning style |
| 5 | Random assignment of technique and control conditions to units | Analysts were randomized into a treatment group using techniques and a control group |
| 6 | Study is a systematic review | Multiple research reports of brainstorming were analyzed using meta-analysis |

To analyze the data, the relation between study quality and reported effect, Rosenthal and Rubin's (1982) binary effect size display (BESD) was used. The purpose of BESD is to determine "the effect on the success rate (e.g., survival rate, cure rate, improvement rate, selection rate, etc.) of the institution of a certain treatment procedure" displayed as the "change in success rate (e.g., survival rate, cure rate, improvement rate, selection rate, etc.) attributable to a certain treatment procedure (Rosenthal and Rubin 1982, p. 166). The BESD is particularly useful for demonstrating in simple terms the impact of an intervention or, in this case, an analytic technique. Effectiveness data was

analyzed by generating a table containing the BESD for each technique. An example is presented below to illustrate how the BESD is displayed with simulated results. In the left-hand column selected techniques are listed and in the middle two columns, reported effectiveness of those techniques. In the far right column, the number of research reports for each technique. For example, in the simulated results, the evidence for red teaming is mixed (50 percent effective, 50 percent ineffective), but this result is based on only 5 research reports.

3.2.2 An Experiment of ACH and Indicators

While the systematic review provided a general snapshot of “what works,” an experiment was also conducted to take a closer look at two techniques, ACH and Indicators. However, developing an experiment for structured analytic techniques needed to be generalizable to intelligence contexts and internally valid. To accomplish this lofty task, graduate security and intelligence students worked on a real-world intelligence task to determine if the techniques improve the rigor and accuracy of analysis.

Study participants were graduate students from the Security and Intelligence Studies (SIS) program at the Graduate School of Public and International Affairs (GSPIA) at the University of Pittsburgh. While it would be ideal to conduct the evaluation with actual analysts, the students provided a useful proxy since they are developing the same core competencies analysts hold, such as domain (e.g. area studies expertise) and

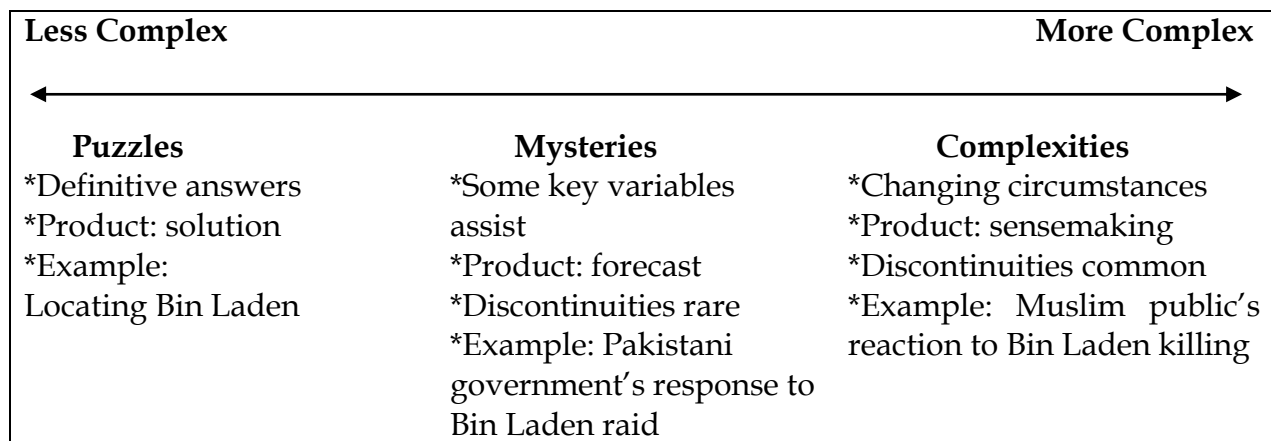
procedural knowledge (e.g. critical thinking skills).²⁵ Students were recruited through three methods: an email on the student list-serv, flyers posted around GSPIA, and class announcements. In total, 21 students volunteered for the experiment and in exchange were provided a small financial incentive (\$20).

To increase the generalizability of the results, an intelligence task was used with a moderate level of complexity and structure: estimating the percentage of chemical weapons the Assad regime would destroy before a UN Resolution deadline. Treverton's (2008) conceptual framework is instructive for classifying this experimental task (see Figure 3.4). Simple intelligence tasks are 'puzzles' with clear solutions and strategies. Locating where Osama Bin Laden is located, for example, has a fairly straight-forward solution strategy (e.g. analyze human intelligence and satellite imagery). 'Mysteries' are complicated by some rare discontinuities, but generally have some key variables with some level of predictability. An example is the Pakistani government's reaction to the Bin Laden raid: the U.S. government probably expected there would be some diplomatic fist shaking, but knew it was unlikely Pakistan would retaliate militarily. Lastly, complexities lack any clear solution strategy consensus. The reaction of potential lone-wolf terrorist to the Bin Laden killing is an example of a complexity because it is difficult to determine how to even begin to answer this question. Of the potential hundreds of thousands or

²⁵ According to the GSPIA website, the security and intelligence studies major "prepares students for careers in the security or intelligence fields with various think tanks or intelligence agencies, such as the FBI or CIA." "Major in Security and Intelligence Studies" Graduate School of Public and International Affairs, accessed 7 May, 2015. Available at: <https://www.gspia.pitt.edu/Academics/DegreePrograms/MasterofPublicInternationalAffairs/MajorInSecurityIntelligenceStudies/tabid/95/Default.aspx>

millions of potential radicalized individuals, it is hard to know where to begin looking for the proverbial needle in the haystack. The Syrian chemical weapons task falls in the middle of the continuum closer to the “puzzle” pole because definitive answers are possible (the percentage of weapons destroyed) and key variables are identifiable to assist in the analysis, yet rare discontinuities are possible (e.g. the war escalates and rebels steal chemical munitions).

Figure 3.4: Intelligence Task Complexity



To the increase internal validity of the experiment, a pretest-posttest design was used (Shadish, Cook, and Campbell 2002). Student participants were randomized into two experimental groups: one using ACH and the other using the Indicators and Signposts of Change technique (henceforth: “Indicators”). The Indicators technique requires analysts to work as a team to list “observable events that one would expect to see if a postulated situation is developing” (U.S. Government 2009, p. 12). For example, analyst could list indicators of an upcoming coup, for example (e.g. the presence of rioting, political assassinations). ACH based on the falsification logic of Karl Popper that

knowledge should advance through a process of conjectures and refutations. The use of ACH is fairly straightforward: analysts start by creating a matrix and then insert evidence in the rows and hypotheses in the columns. Next, each piece of evidence is compared with each hypothesis to attempt to determine what Heuer refers to as “diagnosticity” – the extent to which each of piece of evidence is consistent (or inconsistent) with each hypothesis.

Both experimental groups participated in the experiment during the final week of March 2014 and first week of April 2014 to minimize any advantage one group might have as the destruction process progressed. One week prior to the start of the main experiment, a pilot test was run to check the experimental procedures and instruments with three doctoral students. The Indicators groups, 3 in total with 3-4 students in each, participated in the experiment during the final week of March. Student participants reported to a small conference room and were read the informed consent script as per the Institutional Review Board requirements. Next, each participant was given three documents: the UN chemical weapons resolution, a map of Syria showing chemical weapons sites and areas controlled by the rebels, and a timeline of the weapon disposal process. Participants were given 5-10 minutes to review these documents and then provided a judgment sheet and cognitive reasoning style test. Upon completion of the judgment sheets, participants were led through the two steps of the Indicators technique exercise.

Students first provided possible indicators that might affect the speed at which weapons were removed and destroyed (see a truncated example below in Table 3.6).

Second, each indicator was assessed on 3 point scale for importance on the destruction process (1=very important, 2=somewhat important, 3=not important) and likelihood of the indicator occurring (1=very likely/happening now, 2=somewhat likely, and 3=not likely). For example, one group determined that a major earthquake would have a moderate impact on the destruction process, but determined such an event was unlikely. All of this information was recorded on a whiteboard. Upon completion of the technique, participants were provided with a blank version of the same judgment sheet they completed at the beginning of the exercise. To avoid biasing the answers of other groups, participants were asked to not provide any information or their answers to their colleagues. The entire process required approximately three hours.

Table 3.6: Truncated Indicators Grid

| Indicator | Importance | | Likelihood | |
|---|----------------|----------------|----------------|----------------|
| | Participant #1 | Participant #2 | Participant #1 | Participant #2 |
| Anti-regime forces blocking routes out of country | 2 | 2 | 2 | 3 |
| OPCW can't find sites | 3 | 3 | 3 | 3 |
| Acts of God (weather/natural disaster) | 2 | 2 | 3 | 3 |
| Rebel groups obtain chemical weapons | 1 | 2 | 2 | 3 |
| OPCW comes under attack | 2 | 1 | 1 | 1 |
| Escalation of violence makes areas unsafe | 1 | 1 | 1 | 1 |
| OPCW ability to verify and locate all sites | 1 | 1 | 2 | 1 |
| Assad regime may delay | 1 | 1 | 1 | 3 |
| International incident may draw attention away | 3 | 3 | 3 | 3 |
| Assad regime overthrown | 2 | 2 | 3 | 3 |

| | | | | |
|--|---|---|---|---|
| Assad regime becomes totally uncooperative | 1 | 1 | 2 | 3 |
| Another party gets draw into the conflict | 2 | 2 | 3 | 3 |

While the Indicators was relatively simple to implement without training participants, ACH requires training to implement effectively (Heuer and Pherson 2011, p. 315). The average length for instruction for new analysts in the IC to learn a battery of analytic techniques is two weeks (Defense Intelligence Agency 2011), therefore, the assumption was made that several hours are spent learning each technique. Interviews with State Department Bureau of Intelligence and Research confirmed this assumption—each technique requires 2-3 hours of instruction.²⁶ To simulate this level of training and ensure that participants understood how to use ACH, a four hour workshop was conducted. The curriculum included the rationale for the technique and a full length practice example using a digital version of ACH. After the workshop each participant was provided a short quiz to determine how well they understood the rationale and use of ACH. Across all students, the score was 82 percent on the quiz.

The ACH groups—4 in total containing 4 students in each—followed nearly the same procedures as the Indicators group with the exception that a different technique was used. As with the Indicators groups, the ACH groups were provided with the same documents—the UN resolution, timeline, and map—and asked to make estimative judgments. Next, a facilitator led the use of structured ACH, a simplified version of the

²⁶ Interview with INR analysts, Washington, D.C., February 24-26, 2014.

technique that requires analysts to start with the simplest set of hypotheses (Wheaton and Chido 2006). For example, the two working hypotheses in the experiment were either the weapons would be destroyed or not before the June 1st deadline. This form of ACH was used to reduce the amount of time that is required with a standard ACH which allows for expansive set of hypotheses. To conduct the exercise, participants were first asked to generate pieces of evidence that would disconfirm the two hypotheses then asked to determine if pieces of evidence could be combined or eliminated. In Figure 3.5, below, the pieces of evidence are listed along the left side of the matrix. Next, each piece of evidence was assessed for whether it had a positive or negative bearing on the hypotheses. Where there was disagreement over a simple vote was taken and a flag placed to indicate disagreement, visible as on small flag on the right of Figure 3.5. Then students were asked to consider the sensitivity of each piece of evidence to consider what effect changing a piece of evidence might have on the hypothesis. Lastly, students were provided the judgment sheets and asked to record their answers.

Figure 3.5 An Example ACH Matrix

| | | H: 1 | H: 2 | Ⓜ |
|-----|--|--|--|---|
| | | The chemical weapons agreement will be completed | The chemical weapons agreement will NOT be completed | |
| | Weighted Inconsistency Score ⇄ | -11.0 | -0.0 | |
| | Enter Evidence | | | |
| E13 | Russian military aid has increased recently | I | C | ⚑ |
| E12 | Syria has expressed willingness to comply with agreement | N | N | ⚑ |
| E11 | US condemned use of Syrian chemical weapons | N | N | ⚑ |
| E10 | The perceived costs of destroying weapons outweigh the perceived benefits | II | C | |
| E9 | Willingness to use chemical weapons signals reluctance to destroy them | I | C | |
| E8 | Chemical weapons take an extra 90 days to destroy , rendering the proposal unrealistic | I | CC | ⚑ |

Accuracy was determined by using UN and OPCW reports on the destruction process. The final update before June 30th 92 percent of the weapons were destroyed or

removed (Organisation for the Prohibition of Chemical Weapons 2014). Therefore, judgments closer to the 92 percent mark were considered more accurate. A factorial analysis of variance (ANOVA) was used to compare the judgment means between the experimental groups to determine if there was a statically significant difference. The ANOVA tests was also used to determine if there was any significant change in judgments before and after using a technique. While robust to the normality assumption, ANOVA requires that variances be relatively equal between groups--homogeneity of variance assumption. To test for this assumption, a Levene's test was used and it was determined for the comparisons between groups (Levene Statistic= .054, Sig.=.947) and within groups, including the comparison of pretest and posttests of ACH (Levene Statistic= .238, Sig.=.629) and Indicators (Levene Statistic= .160, Sig.=.695), the homogeneity of assumption was not violated.

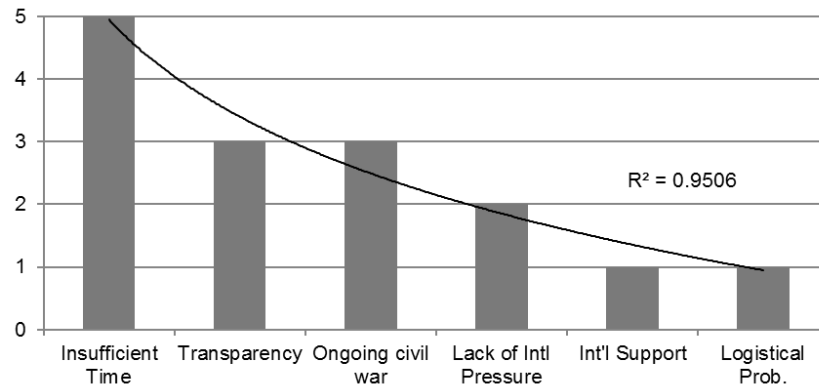
Rigor was determined by coding the narratives of the study participants to determine how well each experimental group explored hypotheses (Miller, *et al.* 2006; Zelik, *et al.* 2007). In the example below (Figure 3.6), two hypotheses were stated that could affect the implementation of chemical weapons agreement (both highlighted), the former to speaking to the rationality of Assad and the latter to the perceived calculus of destroying the weapons. Unique hypotheses were coded for participant and aggregated for each experimental group.

Figure 3.6: Snippet of Narrative with Hypotheses

Assad has shown his willingness to use chemical weapons, and the consequences thus far have not outweighed the benefits.

Next the hypotheses were enumerated and combined for each group and put on a distribution (see Figure 3.7).) and a logarithmic curve fitted to the distribution. The logarithmic curve approximates a distribution of hypotheses that fits most knowledge systems where the ratio of trust to doubt favors the former (Campbell 1977, Dunn 2001, p. 6). Visually, this means the more positively skewed the distribution to the left the more trust there is in a set of favored hypotheses; the less skewed to the right, the less trust and agreement in the hypotheses. For example, below in Figure 3.8 are the hypotheses for the Indicators participants before using the technique. Before using Indicators, participants' judgments closely fit a logarithmic curve of $R^2 = .950$, suggesting agreement on the main hypothesis, time was insufficient for the removal process and therefore it was unlikely to occur.

Figure 3.7: Example of Distribution of Hypotheses



3.3 CONCLUSION

Evaluating structured analytic techniques is no easy task. The secrecy of the IC and methodological challenges of evaluation creates multiple obstacles. However, through employing multiple methods to build interlocking evidence, it is possible to gauge and draw strong inferences. This chapter laid out five methods to evaluate the two research questions. The implementation question was addressed by combining semi-structured interviews with intelligence experts and a survey including follow-up interviews at an IC agency. Each method has complementary strengths: the interviews provide contextual description and the survey allows for the use of statistical procedures to generalize to the rest of INR.

Answering the effectiveness question required combining a systematic review with an experiment. While the systematic review provided a broad overview of all known research studies on the techniques, the experiment examined particular techniques and mechanisms. Combined, these two methods yield credible evidence on whether the

techniques are effective for improving analysis and under what circumstances. The following chapters put these methods into practice with the aim of determining if the IC's efforts to improve foreign affairs judgment have been successful.

4.0 CHAPTER 4: WHO USES STRUCTURED ANALYTIC TECHNIQUES AND WHY?: A SURVEY IN THE INTELLIGENCE COMMUNITY

An open question is how often analysts actually use the techniques. Anecdotal accounts suggest analysts are reluctant to use the techniques on the job (Moore and Hoffman n.d.; Folker 2000). Scholars and practitioners posit several possible explanations why this might be the case, including the perception of the techniques (Marrin 2007; Heuer and Pherson 2011, p. 337), and demographic characteristics of analysts, such as age (Immerman 2011; Fingar 2008), and training (Moore and Hoffman, n.d.). Yet, the most cited reason why analysts do not use the techniques is time pressure (Heuer 1999, p. 85-86; Folker 2000; Khalsa 2009). As explained in [Chapter 2](#), the techniques are perceived to entail significant costs in time and resources. However, beyond anecdote there is little empirical support for this explanation.

This chapter addresses this gap by presenting the result of a survey of 80 analysts at the State Department's Bureau of Intelligence and Research (INR) and interviews with 20 intelligence analysts and experts. The purpose was to understand: 1) how often analysts use the techniques; and 2) what factors are related to their use. The results of the study confirm anecdotal accounts that analysts do not use the techniques on a regular basis. The more important question, however, is *why* analysts might not use the techniques. In this study, one variable proved most important: whether an analyst

received training or had been exposed to the techniques through previous employment or education. To a lesser extent, perceived effectiveness of the techniques was also correlated. Importantly, there is no correlation between the average length of analysts projects – a measure of time pressure--and their preference to use the techniques. This is an important finding because it goes against common wisdom that time pressure is the main reason analysts do not use the techniques.

The chapter proceeds as follows: in section 4.1 the study hypotheses are re-stated, some background on INR provided, and the descriptive statistics from the survey presented. In section 4.2, the finding that training is the most important variable is presented. In section 4.3, three other variables are explored: perception of the techniques, time pressure, and analyst age. The chapter closes with a re-statement of the central arguments and findings.

4.1 FRAMING THE STUDY: TESTING THE VARIABLES AT STATE DEPARTMENT INR

Exploratory interviews conducted for this research with intelligence experts suggested that there is increasing openness to structured analytic techniques over the last decade. Warren Fishbein, an intelligence expert and facilitator of an early effort to implement the techniques in the 1990s, noted that there is greater recognition in the IC that a purely

expertise-based approach to analysis is insufficient.²⁷ Similarly, a CIA analytic methodologist noted that the daily life of analysts has changed over the last two decades, with a greater willingness to collaborate among analysts and “to do the thinking beforehand and use structured discussion before pen goes to paper.”²⁸ Paul Johnson, former Director of the Center for the Study of Intelligence at the CIA, echoes this change in priorities by noting that what constitutes a “good analyst” has changed from mainly regional and functional expertise to include procedural knowledge of how to use structured analytic techniques.²⁹ Despite these anecdotal accounts, it is not clear to what extent structured analytic techniques are used on the job and what variables affect whether an analyst decides to use them.

4.1.1 Study Hypotheses

Moore and Hoffman(n.d.) and Folker (2000) have identified the role of training as an important variable for explaining the adoption of the techniques

Hypothesis 1: There will be a statistically significant ($p < .05$) and positive relation between whether analysts have training in the techniques and use them on the job.

²⁷ Interview with Warren Fishbein, June 11, 2014.

²⁸ Interview with Warren Fishbein, June 11, 2014.

²⁹ Interview with Paul Johnson, June 9, 2014

The role of belief systems (Weiss and Bucuvalas 1980), in particular, how analysts perceive the usefulness of the techniques (Marrin 2007) was expected to be a determinant of whether analysts use the techniques on the job.

Hypothesis 2: There will be a statistically significant ($p < .05$) and positive relation between how effective analysts perceive the techniques and their use of them on the job.

Another variable is the time pressure an analyst is under. The focus on fast paced production of intelligence is claimed to reduce the ability of the analyst to use the techniques (Johnston 2005; Dixon and McNamara 2008).

Hypothesis 3: There will be a statistically significant ($p < .05$) and negative relation between analysts' average analytic project and their use of the techniques.

The demographic shift in the IC towards a younger workforce has led some observers to note that younger analysts will be more comfortable using structured analytic techniques (Immerman 2011; Fingar 2008).

Hypothesis 4: There will be a statistically significant ($p < .05$) and negative relation between analysts' age and their use of the techniques on the job.

4.1.2 State Department's Bureau of Intelligence and Research

The study hypotheses were tested through a survey and interviews with analysts at one of the oldest agencies in the IC. Originating from World War II, INR was formed from

the wartime intelligence agency, the Office of Special Services' (OSS) analytic division and later transferred to the State Department (incidentally, the clandestine wing of OSS became the institutional basis of today's CIA). Today, the agency's mission is to support U.S. diplomacy through all-source intelligence analysis and serve "as the focal point in the State Department for ensuring policy review of sensitive counterintelligence and law enforcement activities around the world" (U.S. Department of State, n.d.). Even with this important mission, INR is tiny both in manpower and resources compared to larger intelligence agencies. In the words of one INR official, its budget is "decimal dust" compared to larger agencies, such as the NSA and CIA (Rood 2006).

These limitations notwithstanding, INR has received significant attention in the past decade. The most notable example occurred in the lead up to the Iraq War when INR gave the sole dissenting opinion in the IC's overall assessment--called the National Intelligence Estimate (NIE)--of Saddam Hussein's nuclear program. Instead of making a firm positive or negative judgment on the existence of Iraq's program, INR stated there was not enough evidence to draw an inference. While INR's success has drawn further interest, little is known of its analytic practices, a gap this research explored through a survey and interviews at the Bureau.³⁰

³⁰ For example, see: Justin Rood, "Analyze This: Inside the one spy agency that got pre-war intelligence on Iraq--and much else--right." *Washington Monthly*, January/February 2005. To my knowledge the only in-depth study INR's analytic practices is an older study by O'Leary (1974): O'Leary, Michael K., et al. "The Quest for Relevance: Quantitative International Relations Research and Government Foreign Affairs Analysis." *International Studies Quarterly* (1974): 211-237.

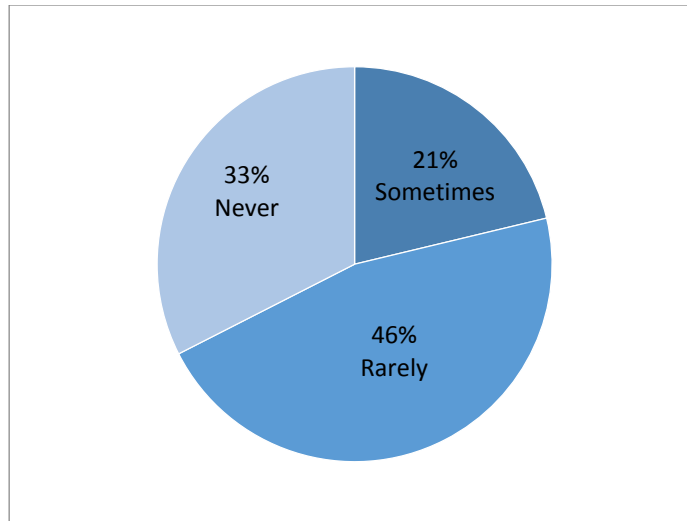
4.1.3 Gauging the Use of Structured Analytic Techniques at INR

The survey responses suggested that structured analytic techniques are not regularly used at INR. In fact, approximately a third of analysts report never using the techniques on the job. The remaining two thirds of analysts are split between those that rarely (33 percent) or sometimes (21 percent) use the techniques (see figure 4.2). The extent to which these numbers reflect the wider IC is only partially known. Interviews with INR analysts suggest that analysts working in larger agencies such as the CIA, DIA, and NSA do use the techniques, although the exact number is not available.³¹

A skeptic could argue that structured analytic techniques are not needed at INR because it is already punching well above its weight without using the techniques. This argument, however, makes an unstated assumption. Specifically, that INR could not improve its performance *further* through using the techniques. Given the promise of the techniques, which is explored in depth in [Chapters 5](#) and [6](#), it is worth exploring what circumstances might be related to whether an analyst chooses to use them. The following sections address these questions at the individual and institutional-level by examining training, demographics, time pressure, and perceptions of the techniques.

Figure 4.1 Use of Structured Analytic Techniques at INR

³¹ Interviews with INR analysts, Washington, D.C., November 2014.



4.2 STUDY HYPOTHESIS 1: ANALYST TRAINING

There is little institutional support for training analysts in structured analytic techniques at INR. As one analyst stated, “INR has no formal training program” and another noted, “I have had training in structured analytic techniques, but not a lot, it was the equivalent of hearing someone talk about it for 30 minutes.”³² As these statements suggest, new INR analysts are not required to take the 4-week “Analysis 101,” a course designed to provide common analytical methods to the IC (Kelly 2007). While the interviews did address why INR is exempted, perhaps one reason is that the agency maintains a certain level of autonomy from the rest of IC.

³² Interviews with INR analysts, Washington, D.C., November 2014.

Given the limited opportunities for INR analysts to learn the techniques, the results of the survey were not surprising: approximately three-quarters of analysts reported having no training. In the place of training in the techniques, INR analysts receive training in the traditional paradigm which the Intelligence Reform Act sought to augment. This training, or as it is better described, “mentoring,” was detailed in Johnston’s (2005) ethnography of the analytic culture of the IC. He found that new analysts were brought into the profession as “journeymen” studying under a “master” with decades of experience (Johnston 2005). During this process the analyst learns in the basic “ins and outs” of their agency with most analytic training focused on analytical writing skills, rather than learning more formal methodologies, such as structured analytic techniques.

Beyond training, INR analysts reported they had limited assistance for using structured analytic techniques. At larger agencies, such as CIA and DIA, there are “tradecraft cells” including a variant called “red cells” that attempt to challenge the prevailing wisdom of analysts.³³

These analytic teams are staffed by methodologists who assist analysts in selecting and implementing methods.³⁴ At INR there is some analytic support in the form of an analytic tradecraft office, although the office is staffed by a single methodologist.³⁵

³³ For one analyst’s experiences working in a red cell unit, see: “Global Agenda: Red Cell Intelligence analyst describes his role as 'devil's advocate' in the CIA,” University of Delaware Daily, April 5, 2012, available at: <http://www.udel.edu/udaily/2012/apr/global-agenda-red-cell-040512.html>

³⁴ Interview with CIA Methodologist, Washington, D.C., June 20, 2014.

³⁵ Interview with INR analyst, Washington, D.C., November 11, 2014.

Analysts interviewed for this study reported that there are other routes for gaining experience with the techniques.³⁶ For example, several analysts reported that interagency discussions with the CIA and the National Intelligence Council (NIC) sometimes involve the use of collaborative techniques to mediate analytic disputes, such as structured Brainstorming and an informal version of Analysis of Competing Hypotheses.³⁷ A couple of analysts received training through previous employment at other agencies. For example, one analyst stated that whether analysts use the techniques depends “on where [the analyst was] educated before they came to INR and on what experience they are bringing to the job.”³⁸ Not surprisingly, the analysts interviewed for this study that came from larger agencies such as the DIA, reported using the techniques on the job at INR because of their previous training.³⁹

Opportunities to get exposure to the techniques through training appears to be strongly related to whether analysts use them on the job. This point was confirmed by the survey data: the relation between whether an analyst reported receiving training and if they use the techniques was highly significant (chi square=13.593, $p=.001$). The high level of significance of the results ($p=.001$) suggested that there is less than a one percent chance that this relation is invalid in a normally-distributed population. Or in other words, the observed results are extremely unlikely to occur by chance alone. In addition, the strength of the relation between training is moderately strong with a Cramér's V value

³⁶ Interviews with INR analysts, Washington, D.C., November 2014.

³⁷ Interviews with INR analysts, Washington, D.C., November 2014.

³⁸ Interview with INR analyst, Washington, D.C., November 10, 2014.

³⁹ Interview with INR analyst, Washington, D.C., November 10, 2014.

of .412. This result is notable because a 0 indicates no association and a 1 perfect association between training and use of the techniques.

The results of the survey suggested exposure to the techniques through training matters and, possibly, the quality of training. Of the 29 INR analysts reporting some training in the techniques, the 20 that reported quality training also reported they use structured analytic techniques on the job; only a single analyst that reported quality training did not report using the techniques on the job. However, caution should be taken in assessing this finding as respondents were not asked to clarify in follow-up interviews what constitutes “quality training.” Future research will need to address this gap and explore quality training.

If training is a key variable at the analyst-level, but what explains the lack adoption or promotion of the techniques at the institutional-level? The short answer is analytic culture. There are two competing, but non-exclusive analytic cultures in the IC: the traditionalist culture focused on deep subject matter expertise and the “generalist” culture focused on the use of explicit methodologies, such as structured analytic techniques. Since the beginning of the analytic reform movement the generalist culture has gained more adherents at larger agencies such as the CIA and NSA. In the generalist culture, analysts are prized for their “organizational flexibility and value in producing current intelligence ...” (Marrin 2013, p. 326). The generalist is able to rotate between research areas focused on specific regions and functional areas (e.g. small arms, terrorism), in theory, applying a “toolbox” of methodologies to different types of analytic

problems. Since 2001, there are more civilian intelligence studies programs built to prepare more generalist analysts for the IC (Crosston and Coulthart, forthcoming).

INR is again, an exception to the rule. While the generalist culture appears to be becoming dominant in other agencies, INR is a stronghold for the traditionalist culture. For example, instead of regular rotations, INR analysts are typically assigned to analyze one region or functional area for the duration of their career. After several years on assignment, INR analysts are expected to gain deep subject matter expertise, if they have not already done so--INR also recruits experts from academia and the private sector (Rood 2006). Adding to the emphasis on subject matter expertise, Foreign Service Officers (FSOs) are rotated through the Bureau to bring firsthand knowledge and area expertise with them. According to one such FSO working at INR: "...by being out there in the field I can begin to read situations... FSOs leaven INR analysis with ground truth."⁴⁰ It is probably due to traditionalist culture that INR eschews analytical training in the techniques which may in turn explain the results of the survey.

4.3 PERCEPTIONS OF THE TECHNIQUES, TIME PRESSURE AND DEMOGRAPHICS

While not as strong as the relation between training, perceptions of the techniques appear to be correlated with adoption (study hypothesis 2). This result speaks to the importance

⁴⁰ Interview with Foreign Service Officer, Washington, D.C., November 25, 2014.

of convincing analysts that there is value in using the techniques. Other variables such as, demographics (hypothesis 3) and time pressure (hypothesis 4) were not supported by the survey data.

4.3.1 Perception of Structured Analytic Techniques

Study results suggested INR analysts perceive the techniques as fairly effective in improving rigor and to a lesser extent, accuracy. Most survey respondents reported that the techniques improved accuracy either a fair amount (17 percent) or a little (60 percent). Interestingly, 20 percent reported that the techniques had no effect on accuracy. However, respondents seemed to have more faith in the techniques for improving rigor. Rigor was described to respondents as the extent to which analysis incorporates multiple hypotheses, viewpoints, and encourages creative thinking (for an extended description of analytic rigor, see [Chapter 2](#)). Among the survey respondents, 32 percent believed the techniques improved analysis “a great deal” and 44 percent “a fair amount.” One analyst echoed the importance of improving rigor through exploring multiple hypotheses by noting that some structured analytic techniques, such as the Analysis of Competing Hypotheses (ACH), would have been useful in the Iraq weapons of mass destruction case.⁴¹ Still, among analysts this view was tempered by the belief that analysis should not be “tool-centric.” A small number of analysts—5 percent—reported the techniques

⁴¹ Interview with INR analyst, Washington, D.C., November 11, 2014.

were not effective in improving rigor, versus the 20 percent that believed they had no effect on accuracy. Two analysts interviewed reported the techniques were a “waste of time” and “not useful,” although these results suggest these analysts are in the minority.⁴²

While not as strong as the relation between training, perceptions of the techniques effect on rigor appears to be correlated with use (chi square=3.83 p=.049). The strength of the relation between perception and use of techniques is not strong, but still notable (Cramér's V= .225). Accuracy was not statistically significant (chi square=1.241, p=.265). These results suggest that there is a relation between perception of the techniques and their use by analysts, although this variable appears to be weaker than an analyst’s exposure to training. These results also speak to an important and obvious point that without a convincing case that analysts ought to try something new, they resist using the techniques (Moore and Hoffman n.d.; Heuer and Pherson 2011, p. 337). Additionally, the responses on perceptions of accuracy suggest analysts have more faith in the techniques to improve the rigor.

Figure 4.2: Perception of Techniques⁴³

| Response | Rigor | Accuracy |
|-----------------|--------------|-----------------|
| A great deal | 5% | 3% |
| A fair amount | 32 | 17 |
| A little | 58 | 60 |
| Not at all | 5 | 20 |

⁴² Interviews with INR analysts, Washington, D.C., November 2014.

⁴³ Survey questions: “To what extent, on average, do you think structured analytic techniques help analysts think in a more effective way (rigor); To what extent, on average, do you think structured analytic techniques help analysts be more accurate or "right" in their analytic judgments? (accuracy)

The INR study presents an interesting puzzle: if most INR analysts believe the techniques are useful, why do they not use them on the job? Two possible explanations emerged in the analyst interviews that might explain the gap between perceived value and adoption: the role of subject matter expertise and problem type. As the discussion above makes clear, traditionalism is the dominant analytic culture at INR. Several analysts reported that their focus on subject matter expertise made structured analytic techniques redundant. “The general belief at INR,” stated one analyst, is that “at this level of expertise you should be doing rigorous thinking that structured analytic techniques make you do.” This view was typically expressed through the idea of thinking “outside the box”: junior analysts were perceived to need help thinking outside the box or they might miss a key piece of information or hypotheses. As another analyst put it, “When you are better versed [in the subject area] you don’t need the techniques to think this way.”⁴⁴ This belief among INR analysts was unexpected because a rationale for using the techniques is to assist all analysts, including subject matter experts (Heuer and Pherson 2010, pp. 5-6). In contrast to the traditionalists’ belief, there is some support from the cognitive psychological literature that experts might not only be susceptible to the same errors as novices, but also that experts make their own unique analytic errors, such as having unwarranted confidence in their judgments.⁴⁵

⁴⁴ Interview with INR analyst, Washington, D.C., November 24, 2014.

⁴⁵ For a discussion of the limits of expertise in intelligence analysis see: Hal R. Arkes and James Kajdasz, “Intuitive Theories of Behavior” in *Intelligence Analysis: Behavioral and Social Scientific Foundations*, Baruch Fischhoff and Cherie Chauvin (eds.), Washington, D.C.: National Academies Press, 2011

Another explanation is that the types of analytic problems that INR analyzes. One analyst noted that the type of problems INR analyzes are qualitatively different from agencies such as the DIA. The analyst noted that the types of analytic problems at these types of agencies appear to be well-structured, especially those dealing with military logistics and planning (“e.g. where the tanks are moving”) versus those at INR dealing with leader’s intentions and beliefs.⁴⁶ Another analyst concurred on this point stating, “linear techniques” which presumably include structured analytic techniques, are “more appropriate for linear problems and sometimes we have linear problems and sometimes we don’t. But my work is a mixed bag—sometimes I can use [the] techniques.”⁴⁷ This belief suggests that some INR analysts are not familiar with the broad range of structured analytics, many of which are useful for diverse tasks including well-structured and ill-structured problems. For example, the Indicators or Signposts of Change technique can assist analysts in structuring problems, which could include identifying drivers that affect a leader’s intentions (see [Chapter 6](#) for an experiment evaluating this technique).

4.3.2 Time Pressure

Since September 11th and the beginning of the “War on Terrorism,” the demand for analysts to produce analysis has increased, while the intelligence production has sped up (Johnston 2005, pp. 26-27; Dixon and McNamara, 2008). However, the survey and

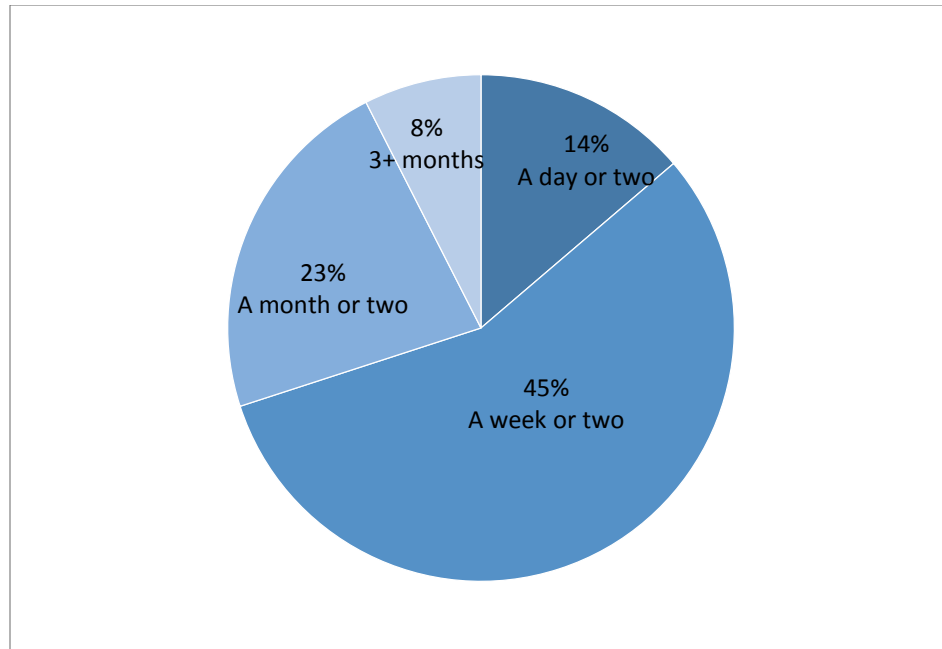
⁴⁶ Interview with INR analyst, Washington, D.C., November 14, 2014.

⁴⁷ Interview with INR analyst, Washington, D.C., November 14, 2014.

interviews from INR suggested that the amount of time pressure, measured by the average project an analyst works on, *is not related to whether they use the techniques*.

In this study time pressure was operationalized by analysts' answer to the question: "On average, how long does it take you to complete an analytic product?" Analysts were asked about their average analytic product rather than asking about how much time pressure they felt, to gain a fairly objective measure of how pressed they are for time. Analytic products can come in the form of presentations, reports, and memos for intelligence consumers. There were 11 analysts under heavy time pressure with their average analytic product taking 1-2 days of work. Of these analysts 7 reported using the techniques. This result is unexpected given the claim busy analysts do not have time to use the techniques. The largest groups were those reporting a moderate amount of time pressure, a week to two weeks for their average product. Of the 45 analysts under this category, 32 reported using techniques versus 13 that do not. As for the 24 analysts that face the least time pressure, a month or more, the results again are not in the expected direction: 15 report using techniques versus 9 that do not. The statistical analysis suggests that there is not a relation between time pressure and the use of structured analytic techniques (chi square= .616, $p=.735$). It is important to note that INR might be different than other agencies in terms of time pressure. One analyst reported that she had been at INR for several years and reported that when she worked at another IC agency the time pressure was considerable. Future research will need to delve into whether the time pressure factor is more important at other agencies.

Figure 4.3: Analysts' Self-Reported Average Analytic Product at INR



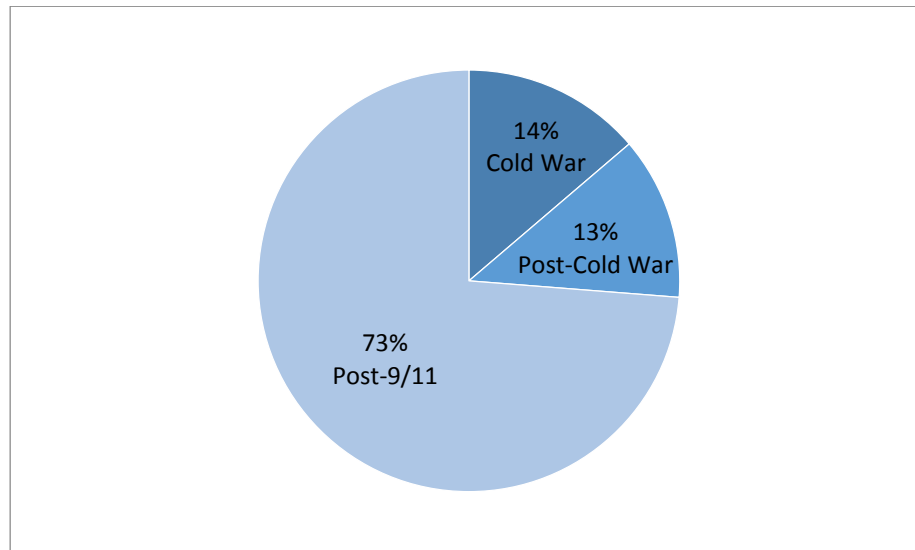
4.3.3 Demographics: Age

Former Deputy Director of National Intelligence for Analysis, Thomas Fingar (2008, p. 24) argues that persuading younger analysts to “adopt new techniques and to work differently than the older generations” is easy. Immerman (2011) echoes Fingar, by noting that ‘Generation Y’ is more “comfortable with, and open to, new techniques that enable collaboration and integration” than previous generations. If these commentators are correct, younger analysts should report using the techniques more than their older colleagues (study hypothesis 3). The results of the survey and interview data tell a different story.

Analysts were asked to report how long they have worked in the IC, as a proxy for their age. Based on the years of service, analysts were divided into three “cohorts”:

the Cold War (pre-1991), post-Cold War (1992-2001), and post-9/11 (2002-onward). For example, an analyst with 25 years in the IC would be counted in the Cold War cohort because they began their career in the late 1980s. The survey results at INR are reflective of this wider demographic shift in the IC; more than half of analysts joined after September 11th (See figure 4.5, below). In fact, almost three quarters of INR analysts surveyed joined INR after the commencement of the War on Terror; 25 percent higher than the average in the rest of the IC (Fingar 2008). The remaining 17 percent of analysts at INR are evenly split between those that joined during Cold War or the interim period between the Cold War and September 11th. In short, if the results are representative, INR is a young agency.

Figure 4.4: Age Cohorts at INR



Of the 59 post-9/11 analysts, 40 report some use of the techniques. In other words, almost two thirds of these analysts report using the techniques. The results for the pre-9/11 analysts are less clear: of the 21 analysts that joined before September 11th, 12 report using the techniques. These data seem to show a weak trend, as post-9/11 analysts do seem to use the techniques more than their older colleagues; however, a chi-square test of the survey results erases any doubt that there is relation between an analyst's age and whether they use the techniques (chi square= .009 p=.924). The Cramer's V score of .011 indicates a very weak correlation, as a 0 indicates no association and a 1 perfect association. The intelligence reformers appear to be wrong, at least as far as INR is concerned.

These results are puzzling: how can age not play a role, especially since newer INR analysts are more likely to be exposed to training, the most important variable identified in this study? The answer seems to loop back to the issue of institutional support: there are few opportunities for young INR analysts to gain familiarity and expertise with the techniques. A young analyst entering INR is unlikely to be very different in analytic training and culture than a senior analyst who joined the Bureau twenty years ago. In short, youth probably matters less at INR than at larger agencies such as the CIA, where the analytic culture has shifted to a more generalist paradigm in the last ten years and techniques are mandated more widely at the institutional level. As one analyst noted, "If

[the techniques] were mandated above and if our leadership made a push the techniques, they would be implemented.”⁴⁸

4.4 SUMMARY OF RESULTS AND CONCLUSIONS

The survey and interviews of INR yielded answers to the research question: How often are structured analytic techniques used and what variables affect their use? In particular, three key findings emerged. First, the survey responses suggested that 1 in 3 analysts do not use structured analytic techniques. While these results can only be cautiously generalized beyond INR, they are the first comprehensive reporting of the implementation of structured analytic techniques.

Second, the main variable that is related to the use of structured analytic techniques at INR is the existence of analytic training. The survey and interviews suggest training is at least moderately correlated with the use of the techniques. Another variable positively correlated with the use of the techniques, although to a lesser degree, was whether analysts perceive the techniques to be effective for improving the rigor of their analysis. Interestingly, while most INR analysts see benefits of using the techniques, many do not use them on the job because either they believe they are more appropriate for novices or consider the techniques inappropriate for their analytic tasks.

⁴⁸ Interview with INR analyst, Washington, D.C., November 14, 2014.

Finally, other variables cited in the literature, such as age cohort and time pressure had no statistically significant relation with the use of the techniques. The latter finding is of great importance because time pressure is the most cited reason why analysts cannot use the techniques on the job (Heuer 1999, p. 85-86; Folker 2000; Khalsa 2009). As an outlier in size, culture, and performance, INR is quite different than other agencies. Therefore, future research will need to determine the extent to which these variables affect the preference to use structured analytic techniques at other, larger agencies such as the CIA and NSA.

5.0 CHAPTER 5: "WHAT WORKS?" A SYSTEMATIC REVIEW OF STRUCTURED ANALYTIC TECHNIQUES

Research on intelligence analysis appears to have changed little in the last several decades. Mangio and Wilkinson (2011, p. 19) discuss the cognitive bias, mirror imaging and conclude that since the 1960s the intelligence literature has discussed it repeatedly. Providing little or no evidence, the research community nevertheless has repeated the basic message: "mirror imaging is bad and an analyst shouldn't do it." A similar situation is true for structured analytic techniques: over the last decade the literature exhorts analysts to use the techniques but provides little or no proof of which techniques are effective and under what circumstances. This observation has not been lost on the wider scientific community. In 2010 and again in 2011, the National Academy of Sciences convened special conferences to assess intelligence and methodologies. The finding of the conferences is clear: "many methods used by or proposed to the Intelligence Community (IC) have not been formally evaluated" (McClelland 2011, p. 95).

As the above quote makes clear, beyond some isolated attempts to examine specific methodologies and issues,⁴⁹ no systematic review of structured analytic

⁴⁹ For an example of validation of intelligence methodology, see: Mandel, David R. "Canadian perspectives: Applied behavioral science in support of intelligence analysis." Invited paper presented at the Public Workshop of the National Research Council Committee on Behavioral and Social Science

techniques has ever been conducted—until now. This chapter presents the results of a systematic review of structured analytic techniques to address this gap. A systematic review is a method to sum up the best research on a specific question. As opposed to a traditional literature review, a systematic review is a transparent method to find, evaluate, and synthesize, the results of relevant research (Campbell Collaboration n.d.). This chapter presents the results of a systematic review of the 12 structured analytic techniques from U.S. Government’s *Analytic Tradecraft Primer* (2009), covering more than 200 studies sampled from a population of thousands of research studies.

The 46 evaluative studies were identified from 261 studies on 6 of the 12 structured analytic techniques. Each of the 46 evaluative studies was assessed for credibility by examining its internal validity using the Maryland Scientific Methods Scale (MSMS) (Sherman 1998) and whether a technique was reported in the study as effective for improving rigor or accuracy. Studies with higher MSMS scores were deemed more credible than those with lower scores. The review found a low credibility evidence base for three techniques—Alternative Futures Analysis, Red Teaming, and Team A/Team B—and no evaluative studies for six techniques (Key Assumptions Check, Quality of Information Check, High-Impact/Low-Probability, Indicators of Signpost/Change, “What If?” Analysis, and Outside-In Thinking). Regrettably, there is low credibility or no evidence for 9 out of the 12 techniques. This gap will need to be addressed by future

research, some initial steps of which are taken in [Chapter 6](#) by evaluating the Indicators of Signpost/Change technique.

Despite the evidentiary gaps in our knowledge of the efficacy of most structured analytic techniques, three techniques did have a robust evidence-base: Devil's Advocacy, Analysis of Competing Hypotheses (ACH) and Brainstorming. Brainstorming, which had a moderately credible evidence-base, was effective in only 40 percent of the studies. Interestingly, collaborative Brainstorming led consistently to a *negative effect* on the quality and quantity of ideas generated. Evaluative studies of ACH suggest it assists analysts in seeking disconfirming evidence, although it did not assist analysts in properly *weighting* disconfirming evidence across three high credibility studies. Although many studies suggest there is no link between seeking disconfirming evidence and judgment accuracy, there is a possible link between evidence weighting and judgment accuracy. This result is consistent with previous research on the importance of weighting and updating beliefs for improving foreign affairs judgment accuracy (Tetlock 2005, p. 217). Devil's Advocacy was found to be effective in challenging and validating assumptions and improving accuracy over consensus-seeking groups, especially in tasks where groups must select one optimal solution.

Section 5.1, presents the descriptive results of the systematic review, the sources, subject matter, reported effectiveness, and credibility of the studies analyzed in the review. The second half of 5.1 provides a broad outline of the overall results of the systematic review. Section 5.2 presents findings regarding three techniques with at least moderately strong evidence. These findings are helpful in understanding the conditions

under which the techniques are effective. The chapter closes with a summary of the main findings and their implications.

5.1 OVERVIEW: “WHAT WORKS?”

The systematic review identified 46 evaluative studies of 6 of the 12 techniques: Alternative Futures Analysis, ACH, Brainstorming, Devil’s Advocacy, Red Teaming, and Team A/Team B. After analyzing the studies for a reported effect of the techniques, and assessing the strength of the research design to determine the credibility of the studies, the results suggested there is moderately credible evidence to suggest that the 6 techniques, *in aggregate*, are effective in improving the rigor and accuracy of analysis in just over half of the studies (skip ahead to figure 5.4 for the overall snapshot of “what works”). However, beyond this overall snapshot, the evidence-base for some techniques are not equally credible as some have stronger research designs than others. Three techniques, ACH, Brainstorming, and Devil’s Advocacy, have at least a moderately credible evidence base. Interestingly, there was a moderately negative correlation (-.49) between study credibility and the reported effect size of the technique. In other words, the lower the credibility of the study the more likely the technique will be reported effective.

5.1.1 Selecting and Describing the Evaluative Studies

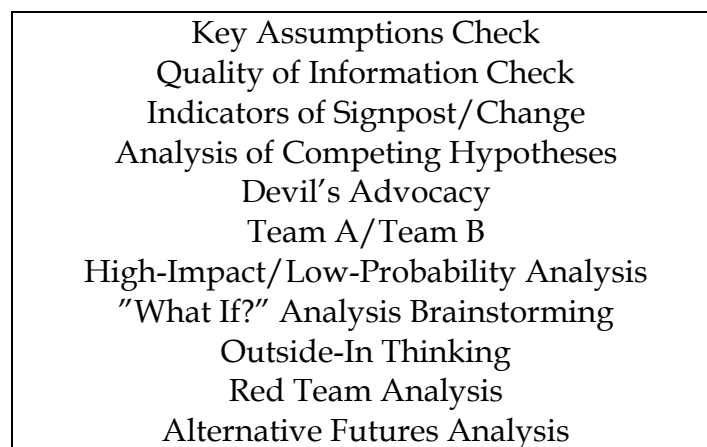
Of the 261 studies sampled (see [Chapter 3](#) for the identification and selection procedures) there were 46 evaluative studies identified. To be included the study had to provide a report, either verbal or numerical of the effect of a technique on analytical rigor or accuracy. Hypothetical reports (e.g. “ACH could be helpful to intelligence analysts...”) were excluded. The remaining studies were excluded for a variety of reasons. For example, 95 studies described how to use a technique in a particular application but provided no explicit evaluation and 36 studies discussed conceptual issues related to using a technique. For a full list of exclusion criteria and procedures, see [Chapter 3](#) and [Methodological Appendix B](#).

The 46 evaluative studies came from diverse sources. In terms of format, most studies came from journals (30) with the remainder from conferences (6), theses/dissertations (5), monographs (4), and other (a blog entry) (1). The focus on journals is not surprising, but the inclusion of studies from non-published, “gray literature” in this review reduces publication bias (Easterbrook 1991), and potentially increases the validity of these results. In terms of subject areas, most evaluative studies came from business and management (18 studies), followed by security studies (7), psychology (7), conservation (6), and other subjects (8).

The review included 12 techniques from the *Tradecraft Primer* (2009) (see figure 5.1, below). Among these techniques there were 6 that had no evaluative research: Key Assumptions Check, Quality of Information Check, High-Impact/Low-Probability,

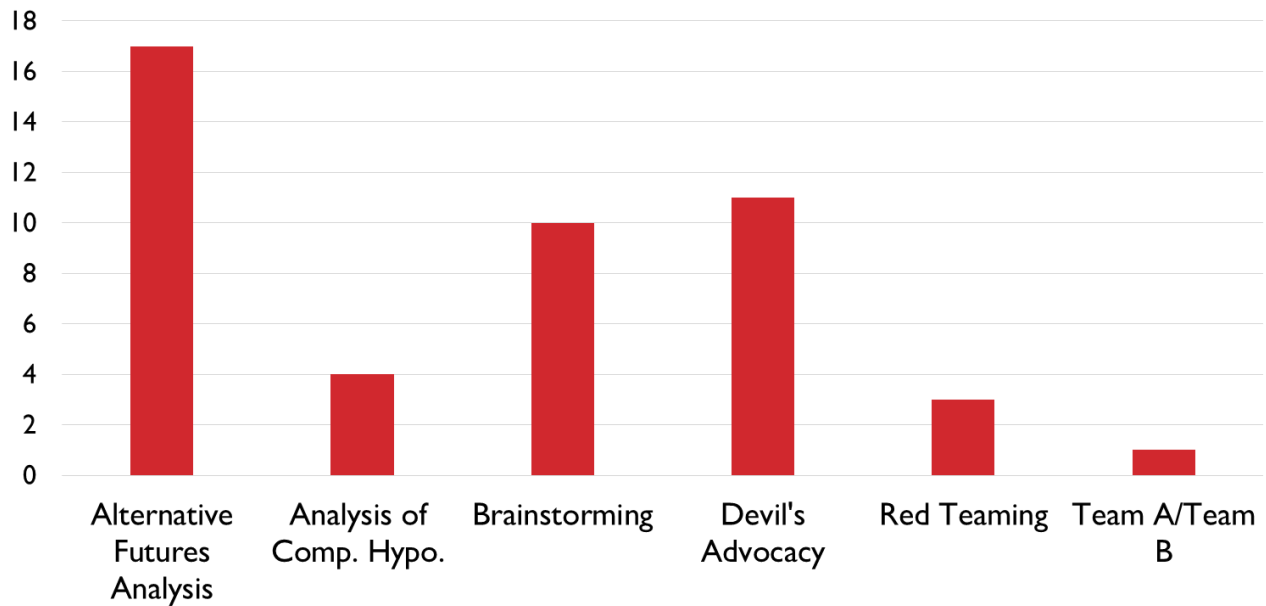
Indicators of Signpost/Change, “What If?” Analysis, and Outside-In Thinking. Upon closer examination, most of these techniques have one trait in common: each was developed specially for the *Tradecraft Primer* (2009) by intelligence trainer, Randy Pherson, and therefore are probably not old enough to have research traditions to produce evaluative studies.

Figure 5.1: Techniques Included in this Study



Other techniques, some from security and intelligence studies, such as ACH, Red Teaming, and Team A/Team B, have a longer history, and therefore, had a few evaluative studies (8 in total). The remaining three techniques—Alternative Futures Analysis, Devil's Advocacy, and Brainstorming—have the longest research traditions and most evaluative studies (38), much of which comes from disciplines outside security and intelligence studies. The only chronological and discernable pattern is that there are decades where more evaluative research was conducted for certain techniques. For example, Brainstorming included 5 studies from psychology from the late 1990s and early 2000s, but after this point there were less evaluative studies.

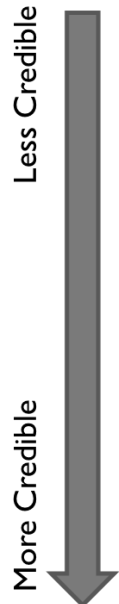
Figure 5.2: Number of Evaluative Studies



5.1.2 Assessing the Credibility of Evidence

From the description above, some techniques have more evaluative research than others, however, not all studies are created equal and vary greatly in terms of the believability of the reported findings. Scholars of methods and research design term this believability as “credibility”; the more credible a study, the greater confidence the intervention (in this case, a technique) had the reported effect (Cook and Campbell 1979; Shadish *et al.* 2002). To assess credibility, a MSMS score was determined for each study (see Figure 5.3).

Figure 5.3: Modified Maryland Scientific Methods Scale



| Level | Description | Example |
|-------|---|---|
| 1 | Correlation between a structured analytic technique and a measure of either rigor or accuracy measures | Analysts reported Brainstorming generates more ideas |
| 2 | Measures of efficacy before and after using a technique, with no comparable control condition | Analysts generated more ideas after using Brainstorming |
| 3 | Measures of efficacy before and after using a technique in experimental and comparable control conditions | Analysts generated more ideas in a treatment group but there was no improvement in efficacy in a control group |
| 4 | Measures of efficacy before and after the program in multiple experimental and control units, controlling for other variables that influence analytic performance | Analysts generated more ideas in a treatment group using Brainstorming but there was no improvement in efficacy in a control group controlling for level of expertise and cognitive reasoning style |
| 5 | Random assignment of technique and control conditions to units | Analysts were randomized into a treatment group using Brainstorming and a control group |
| 6 | Study is a systematic review | Multiple research reports of brainstorming were analyzed using meta-analysis |

Placing the evidence alongside the reported effectiveness of the techniques provides an overall snapshot of “what works” (see figure 5.4). Effectiveness was determined by coding each study according to the reported effectiveness of the technique: 1) the technique had a negative effect; 2) no effect; 3) mixed effect; and 4) positive effect. Techniques that were listed as having a positive effect on rigor or accuracy were considered “effective.” Evidence credibility for each technique was calculated by averaging the MSMS score for all of a technique’s studies and coded into high (5-6), medium (3-4), and low (1-2) after rounding up to the nearest whole number. For example, if a set of studies had an average of MSMS score of 3.8, then it would be rounded to 4 and fall into the moderate credibility range (3-4).

Overall, there is moderately credible evidence for the 6 techniques with evaluative studies (the average is 2.8 on the MSMS scale) (See Figure 5.4). The technique with the strongest evidence base (a MSMS score of 4.5 across all studies) is Devil's Advocacy. Compared to consensus forms of analysis and decision making, Devil's Advocacy is effective more than 70 percent of the time. While a relatively small literature, the ACH studies are highly credible (4.4) and suggest that the technique improves analysis over control groups in half of the reports. The technique appears to partly address confirmation bias but not improve forecast accuracy. Brainstorming had mixed effectiveness and a moderately credible evidence-base (3.8). Nominal or noninteracting brainstorming groups produced more and better quality ideas than collaborative brainstorming groups. The studies from the remaining techniques, Alternative Futures Analysis, Red Teaming, and Team A/Team B, all have low credibility evidence.

Figure 5.4: What Works?: A Display of the Overall Results

| Technique | Effective ⁺ | Ineffective | N | Evidence Credibility |
|-------------------------------|------------------------|-------------|-----------|----------------------|
| Analysis of Competing Hypoth. | 50% | 50% | 4 | High |
| Devil's Advocacy | 72 | 28 | 11 | High |
| Brainstorming | 40 | 60 | 10 | Medium |
| Alternative Futures Analysis | 82 | 18 | 17 | Low |
| Red Teaming | 100 | 0 | 3 | Low |
| Team A/Team B | 0 | 100 | 1 | Low |
| All Techniques | 57% | 43% | 46 | Medium |

*Techniques were considered "effective" if the study reported a positive impact on rigor or analysis

+ "Evidence" is the rounded up mean of all studies' based on the Maryland Scientific Methods Scale scores (1-2= weak; 3-4= moderate; 5-6 strong)

The results of the Alternative Futures Analysis studies are of particular note because while the technique has the most evaluative studies and appears to be highly effective, it had one of the least credible evidence bases. Similarly, Red Teaming appears to also be highly effective in the studies, despite having low credibility evidence. Many researchers including Pillemer and Light (1984, p. 47) have observed similar results in systematic reviews. They have found a negative relation between research quality and the strength of effects, that is, studies with weak designs are associated with high effectiveness. It is worth noting that the Team A/Team B technique also had a weak evidence base although the technique was not found to be effective because partisans used it as a way to push their political views (Mitchell 2006). Researchers have observed similar results in systematic reviews and found a relation between studies with weak designs and reports of either very high or very low effectiveness.

Table 5.1: Binary Effect Size Display of Study Quality and Reported Effectiveness

| | Low | Medium+High | Total |
|-------------|------------|--------------------|--------------|
| Ineffective | 3 (15%) | 16 (64%) | 19 (42%) |
| Effective | 17 (85%) | 9 (36%) | 26 (58%) |
| Total | 20 (100%) | 25 (100%) | 45 (100%) |

A Binary Effect Size Display (BESD) was used to calculate the relation between study quality and effectiveness. The BESD is useful because it visualizes the relation between the two variables in the table and provides a simple correlation. A calculation of

the cells suggests that there is a moderate negative effect of -0.49. This result is notable because the statistic ranges from between -1.0 and + 1.0. A value of 0.0 indicates no association and a value of 1.0 perfect association, whether positive or negative. The implication of this finding is that as long as evaluative designs are weak, we can expect that the techniques will be reported as effective when in fact a stronger design might yield mixed or negative effects. These results paint an overall picture of the evidence for structured analytic techniques, but it is necessary to determine under what conditions the techniques are effective or may lead to negative results. The next section delves into these specific issues.

5.2 UNDER WHAT CONDITIONS ARE THE TECHNIQUES EFFECTIVE?

As the above section detailed, there are 25 evaluative studies with moderate to strong designs; these studies cover 3 techniques. Three findings emerge for creativity, hypothesis testing, and competitive analysis. The four ACH studies suggest how analysts weight evidence might be more important than whether the analyst sought out disconfirming evidence. Another unexpected finding is that Brainstorming in face-to-face groups consistently reduces the quality and quantity of ideas. Instead, it appears that analysts should first brainstorm individually and then combine their ideas. Devil's Advocacy appears to improve the accuracy of judgments and strengthen assumptions, a finding not present in control groups. However, to ensure that the technique is effective,

it appears that analysts should use it in tasks where groups must select one optimal solution from an array of solutions. The next sub-sections describe the research from the moderate and high credibility studies and then provide a synthesis of the main findings.

5.2.1 Analysis of Competing Hypotheses Studies

In total, there were four ACH studies that had at least moderately strong designs.

5.2.1.1 Brant A. Cheikes, Mark J. Brown, Paul E. Lehner, and Leonard Adelman (2004)

In this study, 24 participants from a large research and development corporation volunteered to participate in an experiment evaluating ACH. The experiment was conducted over email. Participants were randomly assigned to an ACH condition and a non-ACH condition. ACH did not reduce availability bias in either group. The authors also examined the distortion effect in confirmation bias (e.g. whether participants misinterpreted evidence as confirming when it should be disconfirming) and a weighting effect of confirmation bias (e.g. giving more importance to support evidence versus providing similar evidence for a non-preferred hypothesis). ACH did assist participants in avoiding the distortion but only participants with intelligence analysis experience avoided the weighting effect.

5.2.1.2 Gregorio Convertino, Dorrit Billman, JP Masur, Peter Pirolli, Jeff Shrager (2006)

The study compared collaborative vs individual (Nominal) use of a digital version of ACH, as well as groups made of members with diverse (Heterogeneous) beliefs vs Homogeneous beliefs. 33 participants were recruited from Stanford University and the Palo Alto Research Center. The results suggest that heterogenous collaborative groups working alone experience a decrease in confirmation bias while using ACH versus homogenous collaborative groups that actually saw confirmation bias accentuated while using ACH.

5.2.1.3 Andrew Brasfield (2009)

70 undergraduate and graduate intelligence studies students attempted to forecast the 2008 Washington State gubernatorial election winner. Study participants were organized into ACH and non-ACH groups controlling for the political affiliation. All groups worked independently for a week and ACH groups used a digitized version of the technique. The ACH group had a slightly higher accuracy at a non-significant level ($P=.421$). The technique was highly effective at addressing confirmation bias ($P=.000$).

5.2.1.4 Kristan Wheaton (2014)

In this study 115 intelligence studies students were assigned to a control group and groups with variations of ACH, including a group without the ability to weight evidence, a group with the ability to weight evidence, and a group with the ability and training to weigh evidence. The groups were given an hour to forecast the winner of a Honduran

presidential election. The most accurate group was the control group followed by the ACH group with training and ability to weigh evidence. Wheaton concludes that “accurate estimates came from analysts who either a) intuitively weighted evidence without the help of a decision tool or b) were instructed how to use the decision tool with special focus on diagnosticity and evidence weighting.”. Also of note, was that “the more accurate groups were also more biased, and while ACH generally helped mitigate bias, it did not improve forecasting accuracy.”

5.2.2 Analysis of Competing Hypotheses Discussion

Since the 1980s, the cognitive biases framework has been one of the main frameworks of intelligence analysis, ushered in through Heuer’s collected essays made public in the late 1990s. As a result of applying the heuristics and biases framework to intelligence analysis, discussion of how to improve analysis typically involves ways of mitigating cognitive biases, in particular, confirmation bias which is “the tendency to seek information...that confirm the tentatively held hypothesis...and *not seek (or discount)* those that support an opposite conclusion... [emphasis added]” (Wickens and Hollands 1999, p. 312). As this definition suggests, there are two properties of confirmation bias: seeking confirming evidence of a favored hypothesis *and* discounting evidence against the analyst’s favored hypothesis.

Results from the systematic review suggest that weighting evidence accurately might be more important than simply seeking disconfirming evidence for a favored

hypothesis. For example, Cheikes *et al.* (2004, p. 15) found that there was little evidence to support the claim that their participants *distorted* negative evidence into positive evidence but rather for *weighting* positive evidence more heavily. Wheaton (2014) found a similar result: his participants were led to disconfirm their hypotheses by ACH but not necessarily to weigh evidence properly. In Cheikes *et al.*'s experiment only those with intelligence analysis experience saw a benefit for weighing evidence properly (but Cheikes *et al.* did not test the accuracy of judgments). Perhaps most importantly, the debiasing of participants in Wheaton and Brasfield's (2009) experiments did not lead to higher forecast accuracy. In fact, in Wheaton's study what mattered most was attaining judgment accuracy was when participants either intuitively weighted evidence or had instruction on how to weight evidence. For example, Wheaton found some participants avoided confirming a favored hypothesis but unless they had additional instruction on weighting evidence, they did not see an increase in accuracy.

The implication of this finding is that analysts should be taught how to weight and assess the credibility of evidence (Wheaton 2014). The importance of weighting sources and evidence confirms Tetlock's (2005, pp. 120-141) finding that the best forecasters are those who update their beliefs. Moving forward for the development of analytic methodologies, researchers and practitioners need to investigate how evidence weighting can be explored to create more valid judgments.

5.2.3 Brainstorming Studies

In total, there were eight Brainstorming studies that had at least moderately strong designs.

5.2.3.1 Henry Lindgren and Fredrica Lindgren (1965)

The study involved 134 university students in a three-phase experiment: in phase 1 participants worked alone without instruction in brainstorming; in phase 2 they brainstormed in groups; and phase 3 again alone. Three judges rated the response for creativity. Intercoder reliability for the judges was high (Kendall's coefficient of concordance was highly significant .76). The results suggest there were significant difference between phase 1 and 2 ($p=.01$). The authors also found no link between culture and the effectiveness of Brainstorming. There was a slight decline in idea quantity and quality from phase 1 to 3.

5.2.3.2 Anne Offner, Thomas Kramer, Joel Winter (1996)

In this study, 180 undergraduate students were randomly selected into nominal (noninteracting) and face-to-face collaborative (interacting). Four variables were manipulated 1) whether a trained facilitator was present; 2) a group recorder; 3) Periodic pauses (interacting groups only); or 4) 5 minute rest periods. The results suggest that brainstorming groups with a facilitator outperformed groups without a facilitator but these groups did not outperform nominal (noninteracting) groups. Other remedies for improving collaborative groups, such as the use of a flip chart, were not effective.

5.2.3.3 René Ziegler, Michael Diehl and Gavin Zijlstra (2000)

In two experiments involving 120 high school and college students, the authors sought to determine the effect of cognitive stimulation, through having group members read others' ideas. To examine this issue, the authors conducted two similar experiments comparing two and four member computer brainstorming groups with and without the opportunity to exchange ideas. The results suggest that computer mediated communication did not result in any increase in creative idea production. These results suggest that using computer mediation might not be effective in improving group collaborative creativity.

5.2.3.4 Sally Blomstrom, F. J. Boster, K. J. Levine, E. M. J. Butler, and S. L. Levine (2000)

207 university students were randomly assigned to one of six experimental conditions resulting in 3 person groups, which included 34 brainstorming and 34 nominal groups. Of the 34 brainstorming groups, 11 were given no training in brainstorming, 11 were given a seven-minute training session, and 12, a 15-minute training session. Of the 34 nominal groups, 12 were assigned to the no-training condition, 10 to the seven-minute training condition, and 12 to the 15-minute training condition. Nominal groups outperformed brainstorming groups in all conditions. Trained groups outperformed untrained groups in terms of ideas generated.

5.2.3.5 Karen Leggett Dugosh, Paul B. Paulus, Evelyn J. Roland, and Huei-Chuan Yang (2000)

To examine the effect of cognitive stimulation, the authors conducted 3 experiments which included more than 200 university students randomized into various treatment conditions, such as hearing ideas of others via audio recordings. The results of the study suggest Brainstormers can be cognitively stimulated as a result of exposure to the ideas of others. Two factors were identified that can influence the effectiveness of this stimulation: 1) the number of ideas a Brainstormer is exposed to and 2) the amount of talking beyond idea expression to which a person is exposed.

5.2.3.6 Henri Barki and Alain Pinsonneault (2001)

This study examined the quality of ideas created through electronic brainstorming (EBS). 96 university student participants were randomized into sixteen 6-member established groups and sixteen 6-member ad hoc groups participated in the study, each randomly assigned to 4 groups: verbal, nominal, EBS anonymous, EBS-non-anonymous. The results of the study suggest that nominal brainstorming groups performed similar if not better than EBS groups. Other variables that were manipulated, such as the effect of trying to “seed” the group with extra ideas, did not have an effect.

5.2.3.7 Nicholas Kohn (2008)

In 3 experiments involving 160 participants, the author found that when participants exchanged ideas in group settings less ideas were explored. In the first experiment the

productivity of nominal (noninteracting) and collaborative brainstorming groups were compared. Collaborative groups explored less categories of ideas than nominal and it appears that the exchange of ideas in collaborative groups led to group conformity. Building off this last point in experiment 2, the author found that conformity increases as the number of ideas a brainstormer is exposed to increases. In experiment 3, the author found that participants systematically exposed to another person's ideas were more likely to conform to the other person's ideas than those who did not receive any exposure. Taking breaks was effective in increasing brainstorming efficiency.

5.2.3.8 Susan Stevens, Courtney Dornburg, Stacey Hendrickson and George Davidson (2008)

The study was an experiment including 69 employees at Sandia Lab working on a real-world "wickedly difficult" challenge. Employees were randomized into one of two groups: collaborative electronic brainstorming or nominal electronic brainstorming. The results of the experiment suggest that "individuals performed at least as well as groups in terms of number of ideas produced and significantly ($p < .02$) outperformed groups in terms of the quality of those ideas." Quality of ideas were those rated by two judges as original, feasible, and effective, in the task.

5.2.4 Brainstorming Discussion

A lesson of the analytic reform movement has been to improve collaboration between agencies and individuals. The message has been so well understood that the IC's website is adorned with the motto: "Collaboration. Commitment. Courage" (Intelligence.gov, n.d.). According to Heuer and Pherson (2011, p. xvi) structured analytic techniques are enablers of collaboration as the techniques prompt "relevant discussion and, typically, this generates more divergent information and more new ideas than any unstructured group process" (p. xvi). However, the results of the systematic review suggest there are circumstances when structured collaboration in the form of face-to-face brainstorming can actually lead to less and lower quality ideas than when analysts work alone *then* collaborate.

Across all 8 studies non-interacting, nominal brainstorming groups consistently generated more and higher quality ideas. Research suggests that when groups engage in face-to-face collaboration they typically struggle as social conformity and pressure limit output (Thompson and Wilson 2014). To address this problem, studies in this review deployed a variety of tactics including, using electronic conferencing, (Dornburg *et al.* 2014; Ziegler, *et al.* 2000), a facilitator (Offner *et al.* 1996), and training (Blomstrom *et al.* 2000), but none of these enabled groups to outperform noninteracting groups in divergent tasks.

The implication of this finding is that if analysts wish to use Brainstorming to produce more and better ideas, they should work independently and then pool ideas.

Attempting to use a face-to-face collaborative group, as suggested in the *Tradecraft Primer* (2009, p.p 27-29) to generate a pool of ideas is unlikely to produce the best outcome, as the 8 evaluative of Brainstorming studies unanimously demonstrate. An upshot of this finding is that intelligence agencies need to rethink how they conduct collaboration. For example, the survey of the State Department's Bureau of Intelligence and Research in the previous chapter suggested that most, if not all, structured brainstorming is done collaboratively.

5.2.5 Devil's Advocacy Studies

In total, there were nine Devil's Advocacy studies that had at least moderately strong designs.

5.2.5.1 Charles R. Schwenk (1984)

In this study four methods were tested, one based on expertise, and three dialectical methods, including Devil's Advocacy. Study participants performed a financial prediction task. The results of the study suggest the dialectical methods were superior when the assumptions in the experimental task were inaccurate. The author also tested how task involvement affect performance of each method and found that greater involvement by participants led the dialectical methods to outperform the Devil's Advocacy and expertise methods.

5.2.5.2 David Schweiger, William Sandberg, and James Ragan (1986)

In the study, 120 MBA students were randomly assigned to four-person groups and randomly assigned to 3 methods groups: dialectical inquirer, Devil's Advocacy, and consensus. The dialectical inquirer is similar to Devil's Advocacy in that it seeks to harness conflict, but through a different procedure. Each group was tasked with analyzing a business management task. Groups were rated on the number of assumptions explored, quality assumptions, quality of recommendations, and a number of other criteria. Where appropriate these criteria were ranked by judges and an acceptable level of intercoder reliability determined. The results of the study suggest that dialectical inquiry and Devil's Advocacy led to higher quality recommendations and assumptions than the consensus method.

5.2.5.3 William Sandberg and Paula Rechner (1988)

120 middle-managers from a Fortune 500 company were randomly assigned to 3 methods groups: dialectical inquirer, Devil's Advocacy, and consensus. Each of the groups was tasked with two business management tasks. The results of the experiment were almost identical to Schweiger *et al.* (1986) experiment: dialectical inquirer and Devil's Advocacy led to higher quality recommendations and assumptions than the consensus method. An additional finding was that group experience with the technique improved performance; in the second task after gaining experience, groups saw increased quality of decisions.

5.2.5.4 Charles Schwenk (1988)

In the study the authors examined the effect of Devil's Advocacy on escalating commitment. 112 undergraduate business students were randomized into four groups: 1) a group receiving success feedback where they were informed their choices in a scenario led to profit (success feedback); 2) a group informed their choices led to a loss; (failure feedback); 3) a group receiving failure feedback along with a recommendation to keep investing; and 4) a group receiving failure feedback along with a devil's advocate report questioning their assumptions. The group receiving the Devil's Advocacy treatment reduced the effects of escalating commitment but the difference with other groups was marginal ($p = <.10$).

5.2.5.5 Charles Schwenk (1989)

The author conducted a meta-analysis of four studies testing dialectical methods, including Devil's Advocacy: Cosier (1978), Cosier (1980), Schwenk and Cosier (1980), and Schwenk (1982) (note: none of these studies were covered elsewhere in this systematic review). Combined, these studies included 252 study participants. A comparison of the studies suggests that basing decisions on an expert is effective when the experts' assumptions are correct. However, "when the assumptions of the expert are not correct, the conflict introduced by both the DA and DI improves decision-making."

5.2.5.6 Audrey Murrell, Alice Stewart, and Brent Engel (1993)

In this study the authors sought to understand how task type affected the effectiveness of Devil's Advocacy. The study considered three types of tasks 1) additive, where group performance is determined by the aggregation of individual effort of all group members; 2) disjunctive, where the group must select one optimal solution from an array of solutions championed by individual group members; and 3) conjunctive, where performance of the group depends on the individual contributions of each group members holding different information. 101 MBA students were randomly assigned to Devil's Advocacy or consensus methods and then assigned an additive, conjunctive, or disjunctive task. In additive tasks consensus approaches are more effective than Devil's Advocacy as the latter retards decision making in this task type. Both methods are equally effective for conjunctive tasks. However, when task structure involves finding a best decision from several alternatives (a disjunctive task), Devil's Advocacy is more effective than consensus methods.

5.2.5.7 Lai Tung (1992)

The author conducted an experiment with 48 groups of 4 members each (192 subjects) to compare a consensus-based approach to two different conflict-based methods, Devil's Advocacy and dialectical inquiry. The results of the study suggest that the conflict-based methods produce more valid assumptions than consensus methods. However, groups using consensus methods perceive their assumptions are stronger than conflict-based methods. The upshot of this result is that while methods like Devil's Advocacy may

produce judgments with stronger assumptions, groups may not believe this about their judgments.

5.2.5.8 M.A. Quaddus, L.L. Tung, L. Chin, P.P. Seow, and G.C. Tan (1998)

In this study the authors started from the assumption the group conflict is productive if channeled properly. To examine this dynamic they designed a study with 116 students randomized into a decision conferencing system either using dialectical inquirer, Devil's Advocacy or consensus approaches. The results are mixed as the authors found that between the three groups there was no difference in terms of conflict generation. Also, there were not any differences between groups in terms of the productivity of conflict

5.2.5.9 Lai Tung and Mohammed Quaddus (2001)

The author examined process level variables such as the type and management strategies of conflict, nor the productivity of the conflict resulting from the use of these approaches. The study was an experimental design consisting of 37 groups with 5 members randomized into dialectical inquiry, Devil's Advocacy or consensus approaches. It was found that Devil's Advocacy increased productive forms of conflict (issue-based conflict) over the other two methods. There was no difference between groups in terms of producing unhelpful forms of conflict (interpersonal conflict).

5.2.6 Devil's Advocacy Discussion

Since the 1970s the IC experimented with competitive analysis. Mitchell (2006, p. 145) describes competitive analysis as “exercises that pit analysts against each other in debating contests designed ostensibly to produce a superior intelligence product from the same pool of raw data. The idea is that “estimative processes” can be sharpened when they are driven by the clash of competing ideas in a structured format.” In other words, conflict should lead to superior analysis.

This systematic review uncovered eight evaluative studies that had at least moderately strong research designs for one competitive analysis technique: Devil's Advocacy. Across nearly all of the studies, conflict-based approaches (which includes Devil's Advocacy) outperformed consensus methods in improving judgment accuracy and validating assumptions. The technique might also be able to mitigate escalating commitment (Schwenk 1988), which can be helpful as it could force analysts to consider revising their beliefs in the face of increasing stakes. Task type also mattered for the effectiveness of Devil's Advocacy: when experts have correct assumptions, Devil's Advocacy was not effective (Schwenk 1989). In tasks where groups must select one optimal solution from an array of solutions the technique is probably more effective. However, when group performance is determined by the aggregation of individual effort of all group members, the technique may hamper the decision making process (Murrell *et al.* 1993). One potential problem is that Devil's Advocacy can lead participants to view

the decision process as less satisfying than consensus-based methods, thus leading analysts to avoid using the technique (Tung 1992).

5.3 SUMMARY OF FINDINGS

The systematic review provided both positive and negative results on the effectiveness of structured analytic techniques. In terms of positive results, the techniques were effective in more than half of the studies. In particular, Devil's Advocacy has a strong evidence base and was effective in most applications. However, there is still the question of how effective the remaining 6 techniques are, given that they have not been evaluated. An unexpected finding was the negative relation between study credibility and reported effectiveness; the lower the credibility the greater the reported effect of the technique.

While these results provide a general overview of the evidence on the techniques, specific findings were extracted and synthesized from the most credible research on ACH, Brainstorming, and Devil's Advocacy. If analysts wish to generate more ideas, the results of the Brainstorming studies were unanimous: brainstorm alone then use a facilitator or an aggregation mechanism, such as a facilitator or software program, to combine ideas. These results are also reflected in the wider literature beyond this review examining how face-to-face collaborating groups struggle to be creative. The studies of ACH suggested a new direction for research examining the role of evidence weighting.

ACH was shown to assist analysts in seeking disconfirming evidence but the technique did not address how analysts weighted disconfirming or confirming evidence which seems to be more important for attaining judgment accuracy. Devil's Advocacy was found to be effective in validating assumptions and accuracy compared to groups seeking consensus, especially tasks where groups must select one optimal solution. In tasks where groups must work together and each group member contributes, however, the technique can obstruct the decision making process by reducing the likelihood that group members will contribute to the analysis (Murrell *et al.* 1993, p. 410). Future research will need to address the gaps exposed by this research. In particular, evaluative research is needed for the 6 techniques with few evaluative studies to understand the conditions under which these techniques are effective.

6.0 CHAPTER 6: AN EXPERIMENT OF ANALYSIS OF COMPETING HYPOTHESES

The results of the systematic review in the previous chapter point to findings of “what works” in intelligence analysis. One key finding is that face-to-face collaboration can limit the number and quality of ideas generated by a group due to social pressure and conformity. Also, the research on Analysis of Competing Hypotheses (ACH) was equivocal; it is not clear if encouraging analysts to disconfirm their favored hypotheses will improve judgment accuracy. There are other areas where there are significant knowledge gaps, including the lack of evaluative research on 6 of the 12 techniques and research on how cognitive reasoning style may interact with the use of the techniques.

This chapter tests the findings from the systematic review on collaboration and ACH and expands the knowledge base to include an ancillary technique called Indicators of Signpost/Change technique (henceforth: “Indicators”) developed during the Cold War for strategic warning (Grabo 2002). The chapter also covers the role of cognitive reasoning styles. Gaps of this kind were addressed using an experiment with 21 graduate intelligence and security studies students randomized to roughly equal-sized groups using ACH or Indicators. The participants made estimates of the percentage of chemical weapons destroyed by the Syrian government as per the requirements of the United Nations Security Council Resolution 2118.

The results of the study partially confirmed the finding that face-to-face collaboration limits idea quantity: the ACH group saw a decrease in the quantity of their ideas and Indicators group saw no change. The importance of this finding is that without a sufficient exploration of hypotheses, analysts will not be able to triangulate their judgments as the theory laid in [Chapter 2](#) suggests. Another interesting finding is that there was no relation found between confirmation bias—measured by the certainty participants had in their hypotheses-- and judgment accuracy. This result supports findings from Brasfield (2009) and Wheaton's (2014) studies. Results from the evaluation of Indicators were disappointing as the technique did not improve the rigor or accuracy of analysis. The results of this study suggest there is no interaction effect between using structured analytic techniques as represented by ACH and Indicators, with cognitive reasoning style. All study participants, regardless of cognitive reasoning style, were equally affected by the two techniques.

In section 6.1, the experimental task and study hypotheses are described. In section 6.2, the results of the study are presented with a focus on testing the study hypotheses. Section 6.3 moves into more a more in-depth discussion of the results and implications for practice and future research. The chapter closes with a re-statement of the central arguments and conclusions.

6.1 FRAMING THE STUDY

The study tested four hypotheses using an experiment with 21 graduate security and intelligence studies in a simulated intelligence task.

6.1.1 Task Background and Procedures

In the wake of the Arab Spring in 2011, Syria was rocked by internal violence between the authoritarian Assad regime and a mosaic of competing rebel groups. The resulting power struggle is fueling one of the bloodiest conflicts of the 21st century with an estimated 76,000 killed in 2014 alone (Gladstone 2015). In August, 2013, the Assad regime used rockets tipped with nerve gas against an opposition neighborhood in Damascus, Syria's capital. Casualty estimates vary significantly depending on the source, but range from the US's assessment of approximately 1,500 to a French assessment of 281 killed (Nikitin *et al.* 2013, p. 15). While this was not the first time the Syrian government had used chemical weapons on civilians--it had done so on a smaller scale previously in the year⁵⁰--the scale and visibility of the attack brought intense international attention.

In early September, as the US was weighing military options, the Syrian government signaled it was willing to seek diplomatic solutions to the crisis. From these early talks, a framework emerged for eliminating Syria's chemical weapons stockpile by

⁵⁰ For a discussion of previous attacks, see: Mary Beth D Nikitin, Paul K. Kerr, Andrew Feickert. "Syria's Chemical Weapons: Issues for Congress." *Congressional Research Service*, September 30, 2013, pp- 11-15, available at: <http://www.fas.org/sgp/crs/nuke/R42848.pdf>

joining the Chemical Weapons Convention and working with the Organisation for the Prohibition of Chemical Weapons (OPCW). In late September, United Nations Security Council passed Resolution 2118 setting a target of mid-2014 for the removal and destruction of Syria's chemical munitions.

In mid-March 2014, an open call for this research was made for study participants at University of Pittsburgh's Graduate School of Public and International Affairs (GSPIA). Participants were offered a small financial incentive (\$20). Most of the participants were first year students (13) and the remainder second year (8). The average age of participants was 25, although 5 did not report their ages. These results are similar to those from the INR study and broadly representative of the demographic profile of the IC towards a younger workforce (Immerman 2011; Fingar 2008). Approximately 60 percent of the study participants were male which is almost identical to a 2009 estimate of the IC a whole and the results of the survey at INR (47 men and 33 women filled out the survey). In sum, the demographic and skill profile of the participants are similar to IC analysts.

The 21 participants were randomized into one of the 3 groups using ACH or 3 groups using Indicators. All groups conducted a three hour analysis session with a facilitator to make judgments regarding Syria's ability to comply with the destruction schedule and fill out a cognitive reasoning style questionnaire. Each participant was provided a backgrounder on the Syrian chemical weapons agreement and judgment sheets both at the start of the analysis session and after. ACH groups received three hours of training on the technique prior to the analysis session.

6.1.2 Study Hypotheses

The experiment addressed four hypotheses:

Hypothesis 1: Face-to-face group collaboration will reduce the number of hypotheses considered by each group

The review of the medium and high credibility Brainstorming studies in [Chapter 5](#), suggests that face-to-face collaborative groups limit creativity through social pressure and self-censorship. This finding is also echoed in diverse literatures, such as social psychology (Thompson and Wilson 2014), forecasting (Armstrong 2006), and was recognized in early efforts to develop forecasting techniques (Rescher 1998). Therefore, it was expected that the hypotheses participants generate in aggregate before collaborating, will be greater than after the number of hypotheses generated after face-to-face collaboration, regardless of the technique used. This hypothesis is particularly important for ACH because the technique calls for analysts to collaborate in order to generate a full set of plausible hypotheses (Heuer and Pherson, 2011, p. 32).

Hypothesis 2: ACH will decrease certainty of the rival hypotheses

According to the *Tradecraft Primer* (U.S. Government 2009, p. 14), “Analysts often are susceptible to being unduly influenced by a first impression, based on incomplete data, an existing analytic line, or a single explanation that seems to fit well enough.” ACH is designed to mitigate this effect and force analysts to attend to multiple rival hypotheses, thereby decreasing their certainty in a single or small set of hypotheses.

Indeed, the systematic review in [Chapter 5](#) provided some evidence that ACH addresses this problem (Brasfield 2009; Wheaton 2014). Therefore, it was expected that after using ACH participants should have less certainty in a single or a few hypotheses.

Hypothesis 3: Increased uncertainty of multiple rival hypotheses will not be related to improved judgment accuracy

Brasfield (2009) and Wheaton's (2014) studies of ACH suggest that leading analysts to disconfirm their favored hypothesis may not result in more accurate judgments.

Hypothesis 4: An open cognitive reasoning style will be positively correlated with hypothesis exploration after using ACH and Indicators

Tetlock (2005) found that study participants with an open cognitive reasoning style were overwhelmed by many rival hypotheses after using Alternative Futures Analysis. On the other hand, those with more closed cognitive reasoning styles were not likely to consider new hypotheses after using the technique. A similar question is whether Indicators or ACH will result in more consideration of more rival hypotheses by those with open styles versus those with closed styles.

6.2 TESTING THE STUDY HYPOTHESES

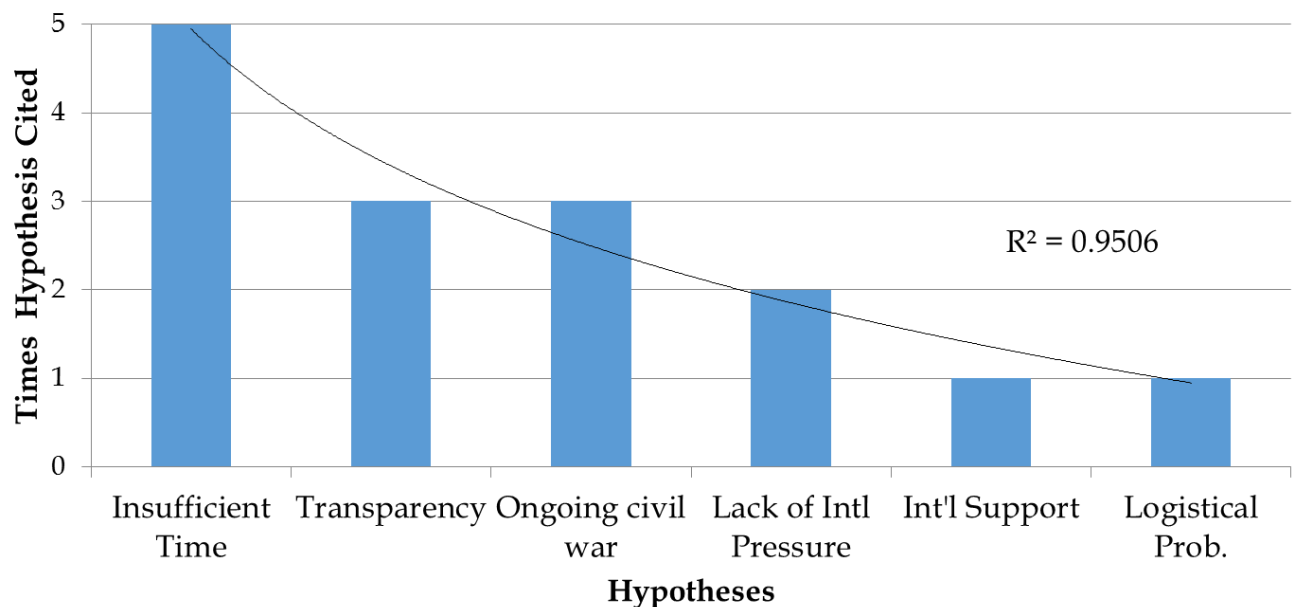
The results of this experiment cohere with many of the findings from the systematic review. There was modest support for study hypothesis 1, that face-to-face collaboration reduces group creativity. While the Indicators group saw no change in the number or quality of hypotheses generated, the ACH group saw a decrease in the number of hypotheses. However, there was more support for study hypothesis 2 as ACH groups did have a reduction certainty surrounding the main hypotheses. The most striking result was that ACH's assistance in encouraging analysts to not focus on a single hypothesis does not improve forecast accuracy. There was no support for study hypothesis 4 that reasoning style interacts with the use of the techniques.

6.2.1 Study Hypotheses 1 and 2: Collaboration and Multiple Rival Hypotheses

The Indicators groups identified 6 rival hypotheses, which included the lack of transparency and ongoing civil war. To measure how much disputation was present, participant's narrative responses were coded for hypotheses, then enumerated, and a logarithmic curve was fitted to the distribution (see [Chapter 3](#) for details). The logarithmic curve approximates a distribution of hypotheses that fits most knowledge systems. Dunn (2001, p. 10) suggests that the conformity of hypotheses to the distribution can be assessed by applying goodness-of-fit procedures with a semi-logarithmic transformation. Before using Indicators, the distribution of participants' judgments

closely fit a right-skew distribution observed in knowledge systems of many types. The goodness-of-fit measure $R^2 = .950$, suggests a high degree of conformity (95 percent) in the group's hypotheses. In other words, before using the Indicators technique, most participants cited the same hypotheses when providing a rationale for their judgments.

Figure 6.1: Indicators group's hypotheses before using the technique

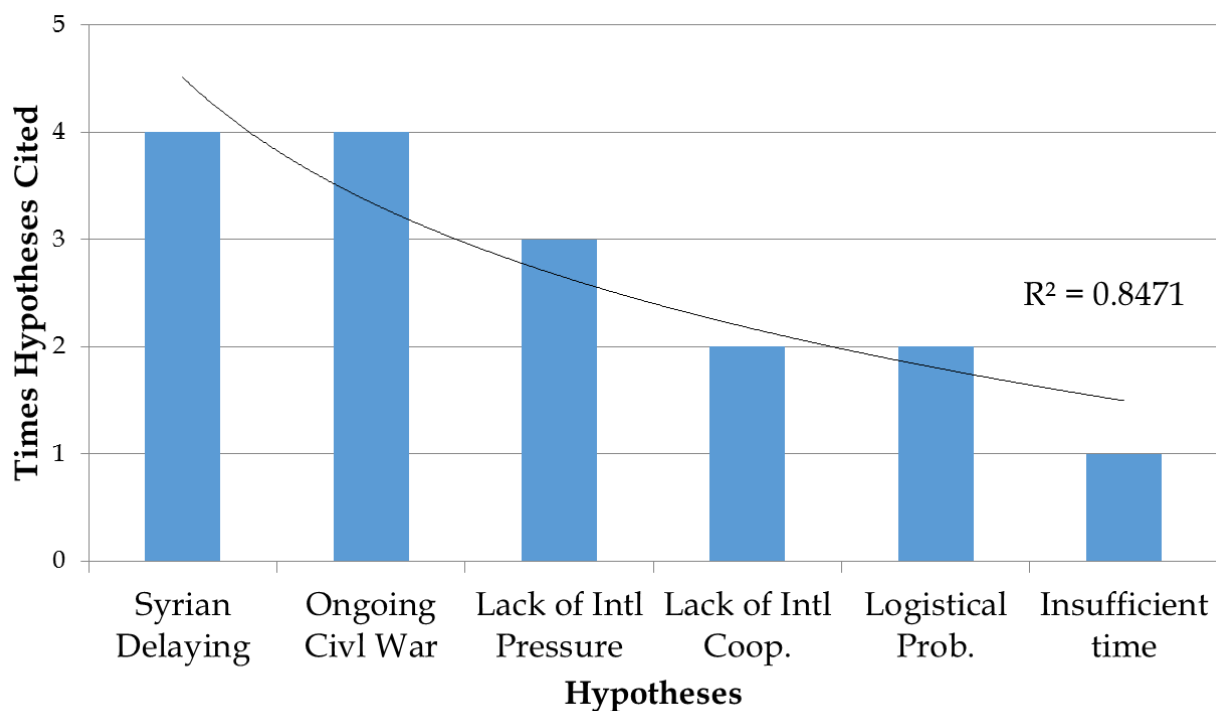


After making their initial judgments, the Indicators participants used the technique in the experimental task and made their judgments once more. Turning to the findings in Figure 6.2, there is mixed support for study hypothesis 1 on the effects of group collaboration. After using the technique, participants did not identify new hypotheses beyond the original six, but did not see a reduction either. Interestingly,

Indicators seems to have slightly increased uncertainty across the distribution of hypotheses which is indicated by the flattening of the distribution and the slight reduction in the right-skew. In particular, there seems to have been more doubt in the “insufficient time” hypothesis. After using the technique this hypothesis went from the most to least cited and it appears that Indicators participants were more inclined to see the Syrian government as blocking or delaying the process to their advantage.

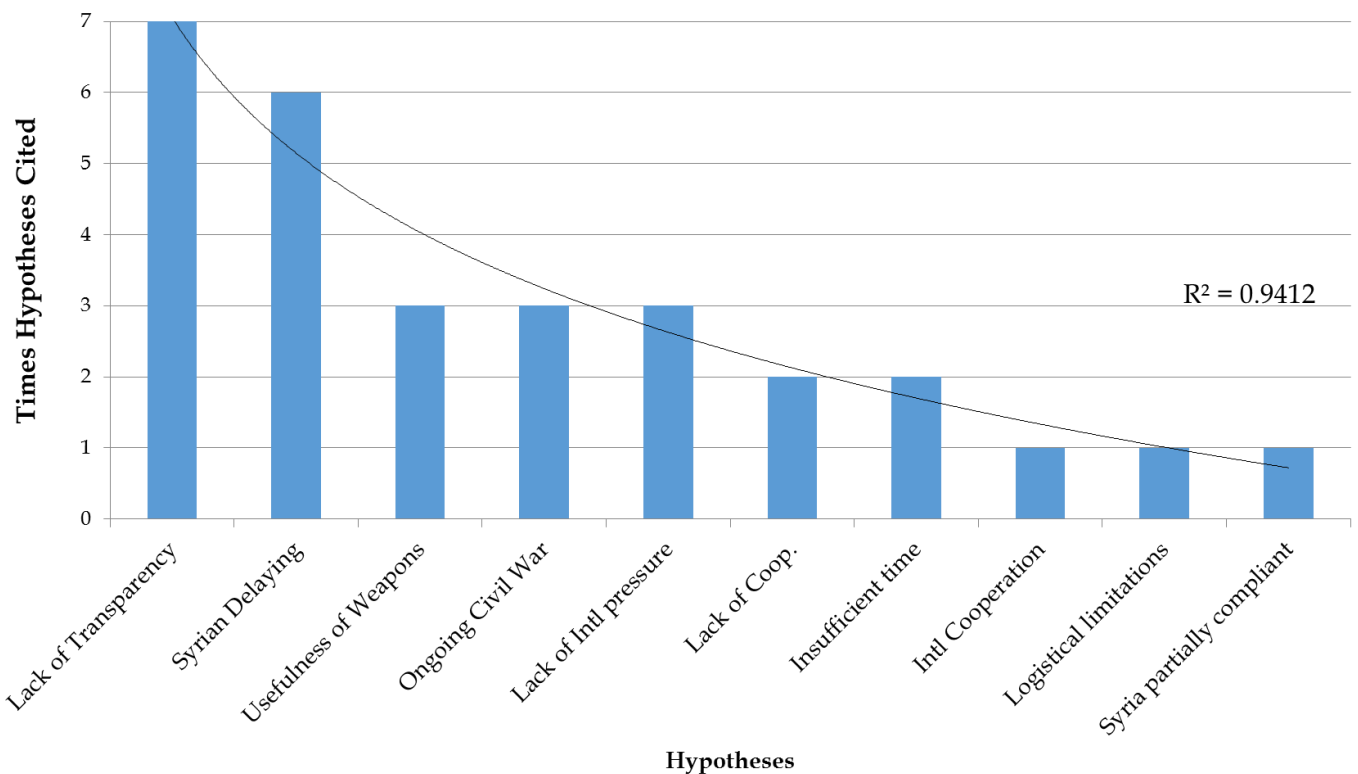
In summary, the Indicators technique appears to have not increased the number of hypotheses considered by the participants and led to a minor decrease in the certainly participants had in the central hypotheses.

Figure 6.2: Indicators group’s hypotheses after using the technique



The ACH results tell a somewhat different story. Before using ACH, participants in this experimental group identified ten rival hypotheses in their judgments (see figure 6.3). As with the pretest of the Indicators group, the hypotheses generated by the ACH groups before using the technique fit a logarithmic curve well with a R^2 value of .941, similar to the .95 value of the Indicators groups. Again, these initial pre-test results suggest certainty in a few rival hypotheses, specifically in the lack of transparency and Syrian delaying being the main hypotheses to explain the outcome of the removal and destruction process.

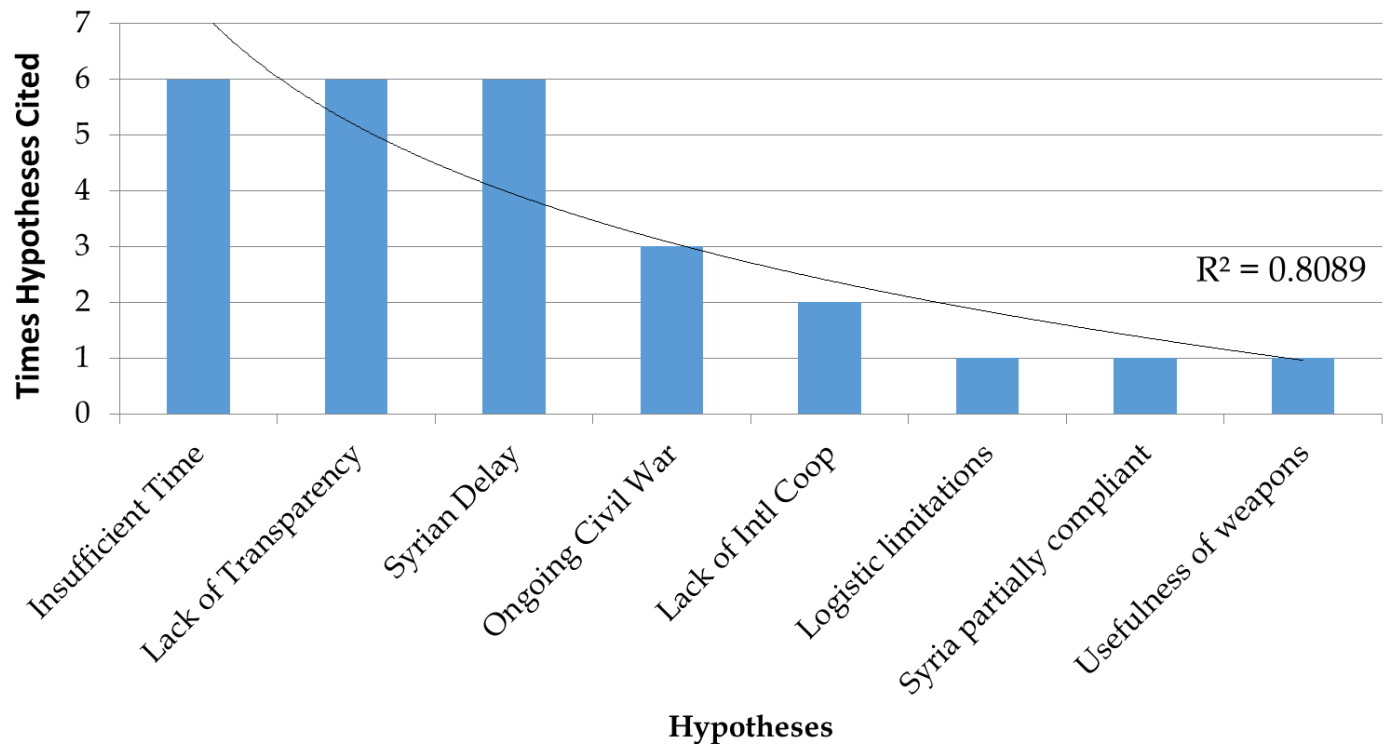
Figure 6.3: ACH group's hypotheses before using the technique



Use of ACH seems to have reduced certainty slightly over Indicators, thus providing some support for study hypothesis 2. After using the technique ACH groups saw a flattening out of hypotheses as the top three hypotheses received equal support from participants (see figure 6.5). The logarithmic curve also fits less well in the posttest judgments, with an R^2 measure shifting from .941 to .808; this change represents a modest, but greater change in pretest to posttest than the Indicators groups ($R^2 = .950$ to $R^2 = .847$). This modest evidence that ACH decreases confirmation of a single favored hypotheses coheres with evidence from Cheikes *et al.* (2004), Brasfield (2009), and Wheaton (2014).

Examining the effect of collaboration on creativity, the results support study hypothesis 1 that face-to-face collaboration reduces creativity. In the pretest condition, participants identified 10 rival hypotheses, however, after using the technique the number dropped to 8. This is not a large decrease, but the omitted hypotheses after face-to-face collaboration might have had an impact on participant's final judgments. For example, the hypothesis that Syria's partial compliance meant that the removal process could be completed by the June 30th deadline disappeared from participants' judgment narratives after using ACH. Greater consideration of this hypothesis might have led participants to estimate the number of weapons to be removed and destroyed to be higher, and perhaps, more accurate.

Figure 6.4: ACH group's hypotheses after using the technique

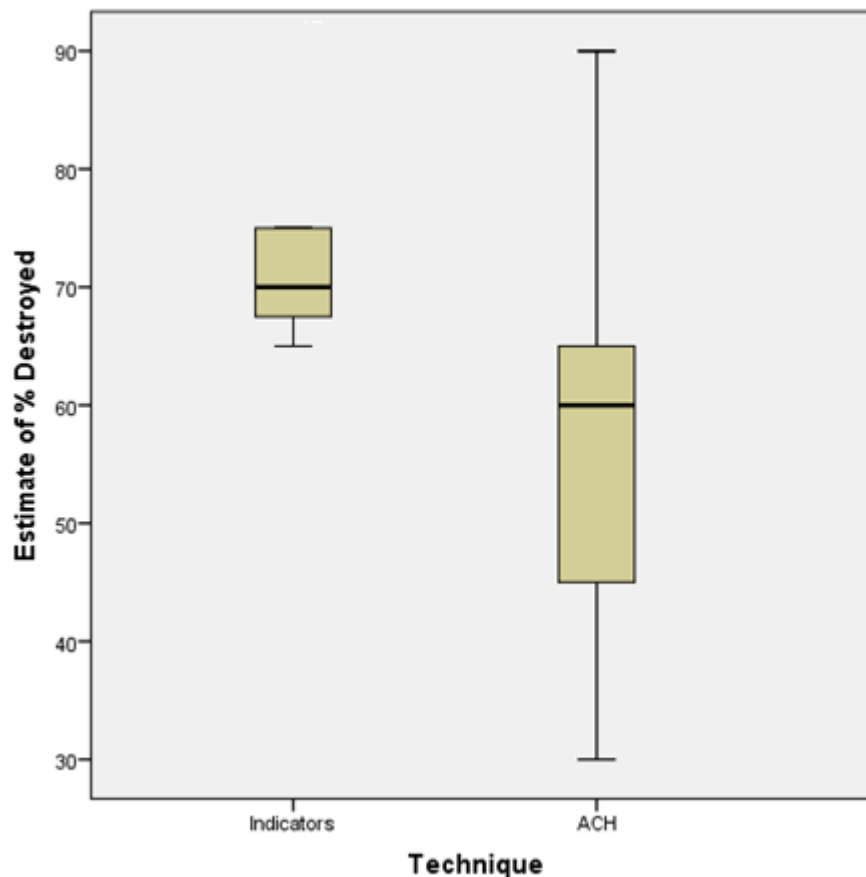


6.2.2 Study Hypothesis 3: Hypothesis Disconfirmation and Accuracy

The final update before June 30th on the chemical weapons removal and destruction process was 92 percent (Organisation for the Prohibition of Chemical Weapons 2014). Therefore, a judgment closer to the 92 percent mark was considered more accurate. Study Hypothesis 3 states that there should be no relation between increasing uncertainty and accuracy. The findings of this experiment bore out this result. In fact, even though the ACH group had less certainty in their hypotheses, their accuracy was slightly lower than the Indicators group. Two studies have found a similar result: assisting analysts to disconfirm their hypotheses does not necessarily improve forecast accuracy (Brasfield 2009; Wheaton 2014).

An unexpected and interesting finding was the greater variation in ACH judgments compared to those from the Indicators groups. The Indicators groups' judgments cluster closely around 67 to 75 percent as indicated by the boxplot in figure 6.6, while ACH judgments had much wider dispersion. One possible explanation for is outcome is that focusing on multiple rival hypotheses among ACH groups led to more variation in their judgments. To test this assertion, it was necessary to inspect the pretest and posttest judgments for both experimental groups.

Figure 6.5: Between Groups Comparison



The pretest and posttest measures of Indicators groups suggest that the technique increased the divergence of judgments, although this difference does not rise to a level of statistical significance ($F=.172$, $p=.685$) (see figure 6.7 below). Indeed, the means changed very little from the pretest to posttest (estimated 72% initially then after using the technique dropped to 68%) suggesting that the Indicators technique had little or no effect on accuracy. ACH participants also did not see a statistically significant difference between pretest and posttest judgments ($F=.265$, $p=.611$) with the means changing little from the pretest to posttest. Participants estimated 57 percent initially, then after using the technique, the average dropped to 55 percent (note: the dark line indicates the median, not the mean judgment) (See figure 6.5). The only noticeable change in the accuracy measure is that the variation in judgments appears to have increased, or in other words, consensus on the number of weapons that would be destroyed seems to have decreased after using the technique. The red lines in figure 6.5 highlight this change below.

Figure 6.6: Pretest and Posttest of Indicators

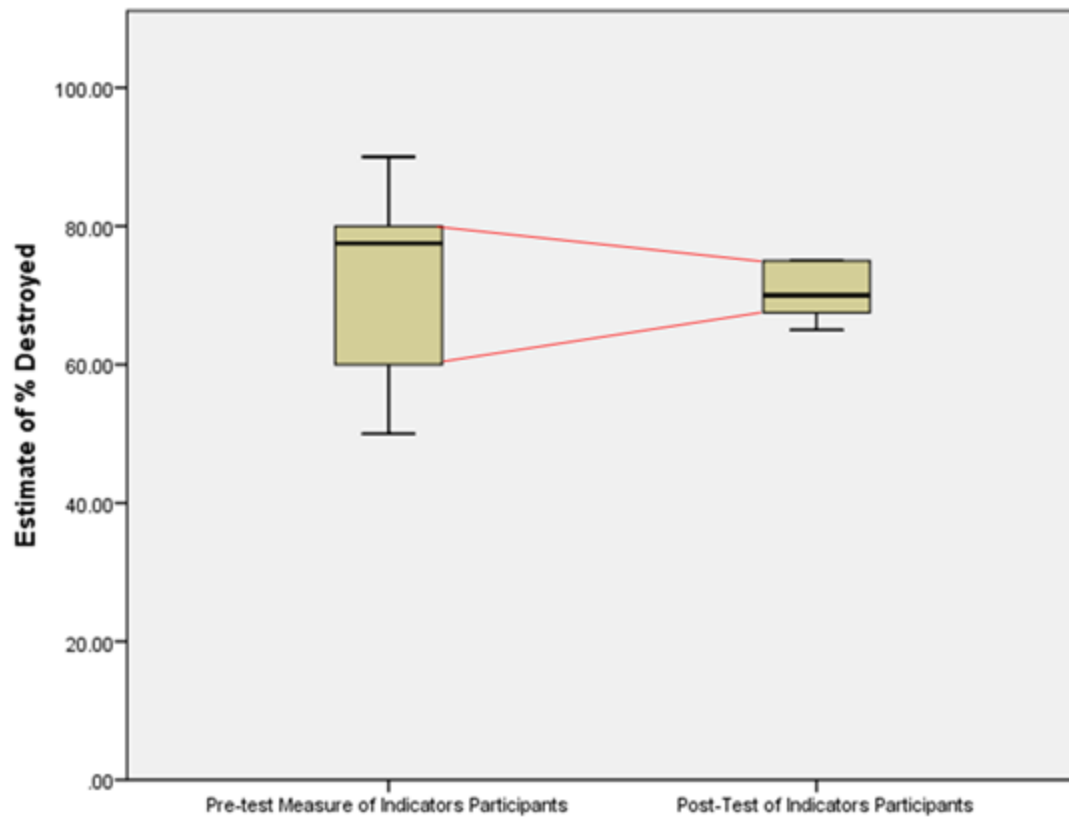
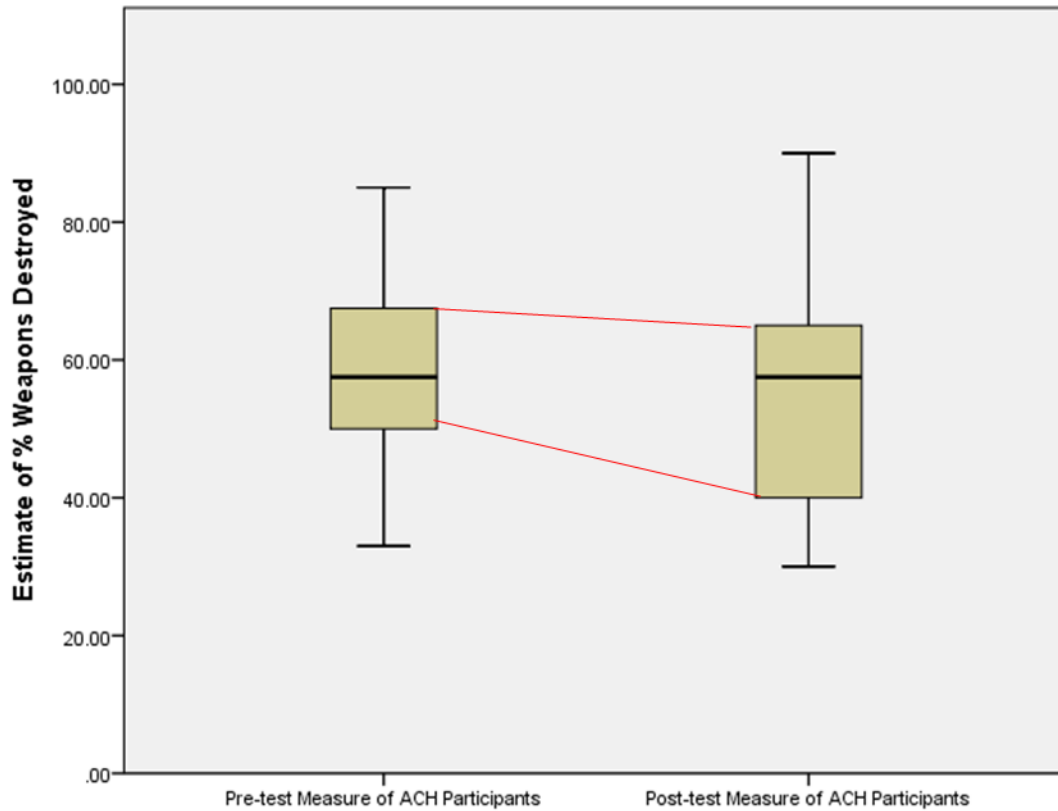


Figure 6.7: Pretest and Posttest of ACH



6.2.3 Study Hypothesis 4: Interaction effect of the techniques and reasoning styles

Given the open cognitive reasoning style of the foxes, it was expected they would identify a greater quantity of hypotheses. To investigate these variables, the participants were divided into those with more open ('foxes') and closed ('hedgehogs') and the quantity and novelty of hypotheses examined. Whether a hypothesis was considered "novel" was determined by where it fell on a distribution of hypotheses generated by foreign affairs experts. As with the student participants, experts were asked to make judgments and provide rationales. Hypotheses were then coded and aggregated on a cumulative

frequency diagram. After the 19th expert the number of rival hypotheses rapidly decreased. This result was predicted by Bradford's law of scattering: after an initial rapid increase in new information (see Figure 6.11). Student participants' hypotheses were considered novel if they fell after the stopping rule.

As expected, foxes identified more hypotheses (11) and novel hypotheses (3). Interestingly, each of the novel hypotheses suggested the completion of the removal and destruction process by the June 30th deadline. For example, one hypothesis was that international support would be strong enough that destruction within the deadline was possible. Another was that Syria has signaled it would cooperate and therefore the process would be completed before the deadline. These novel hypotheses are important given that Syria nearly completed the process by June 30th. In short, the novel hypotheses were positive in suggesting the process would be completed.

Figure 6.8: Foxes hypotheses before using a technique

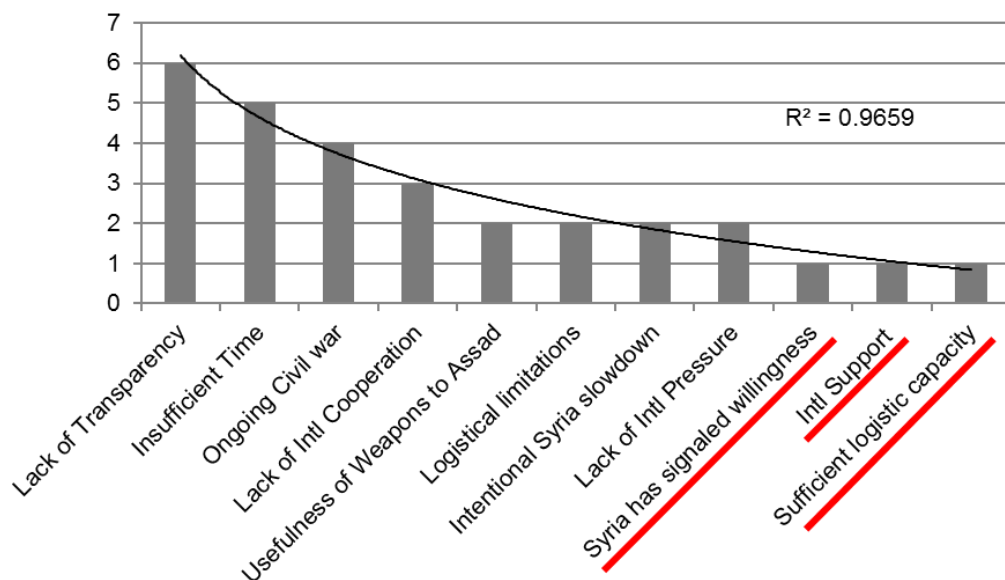
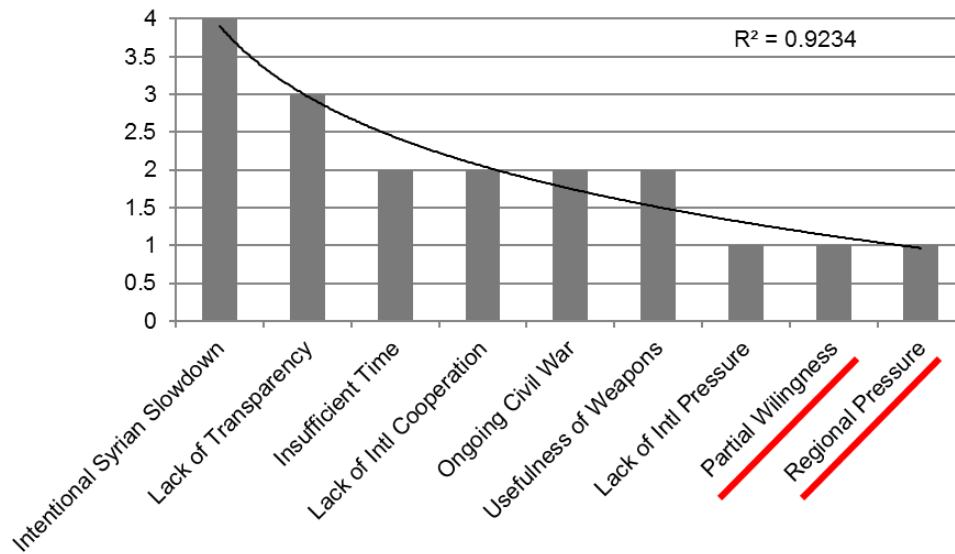


Figure 6.9: Hedgehogs hypotheses before using a technique



After using a technique, the uncommon hypotheses almost completely disappear from the foxes and hedgehogs' judgments with only one remaining. As discussed above, this result could be due to group conformity narrowing the analysis. Both groups also saw a reduction in the number of hypotheses generated. While the hedgehogs began with 9 hypotheses, the count dropped to 7 after using a technique. Both of the hypotheses dropped were novel and related to Syria successfully completing the removal and destruction process. Foxes dropped from 11 to 9 hypotheses and also did not include two novel hypotheses after using the techniques. These results are interesting because they suggest that for both groups, hedgehogs and foxes, the techniques had the effect of narrowing the analysis. These results seem to disconfirm study hypothesis 4 – both types of reasoning styles were equally affected by the techniques.

Figure 6.10: Foxes hypotheses after using a technique

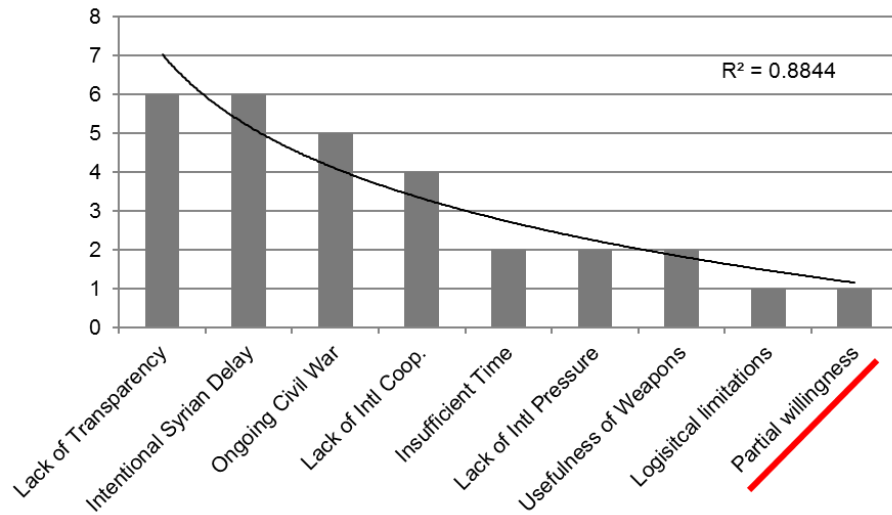
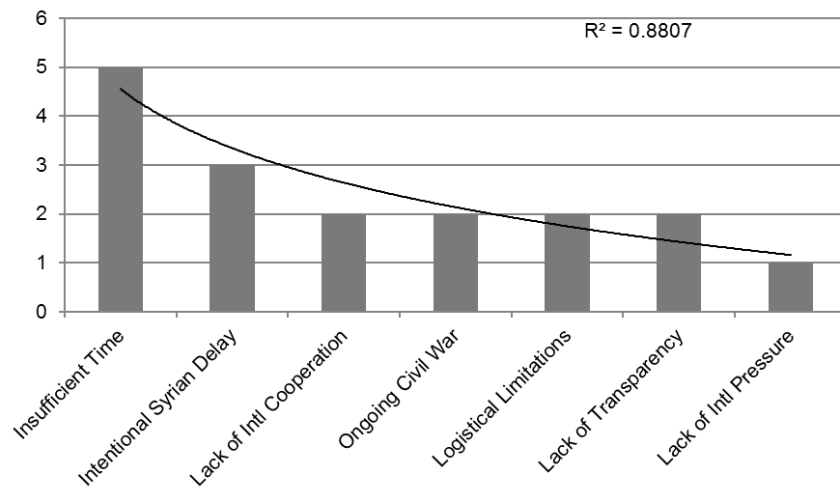


Figure 6.11: Hedgehogs hypotheses after using a technique



6.3 DISCUSSION OF RESULTS

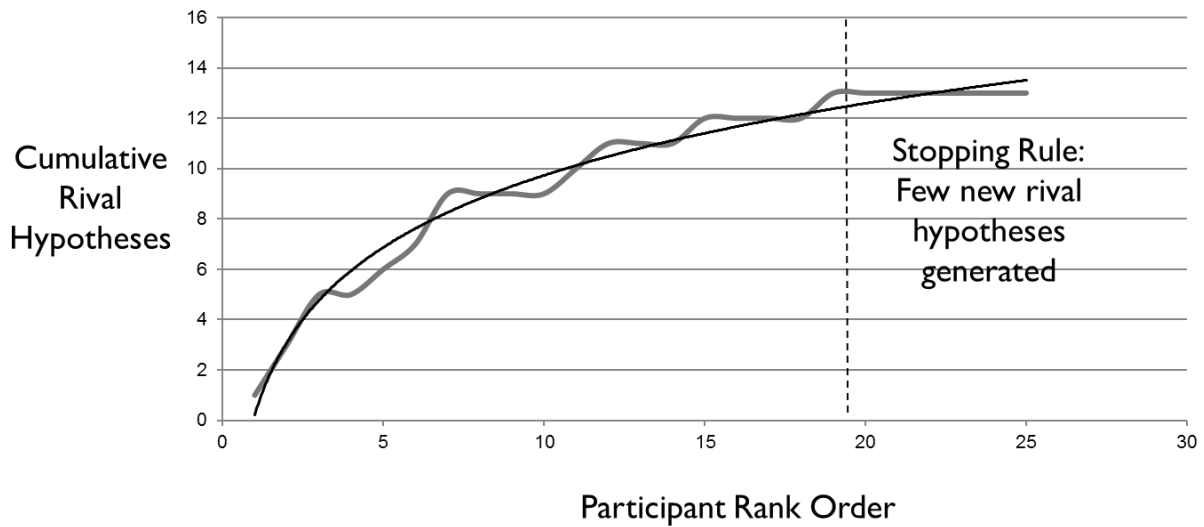
6.3.1 Face-to-Face Collaboration is a Limiter of Creativity and Next Steps

In support of study hypothesis 1, face-to-face collaboration led to no improvement in the quantity of hypotheses generated in both experimental groups. This result coheres with the research on Brainstorming presented in systematic review in [Chapter 5](#) and research from a variety of fields, including social psychology (Thompson and Wilson 2014) and forecasting (Armstrong 2006), and therefore increases the credibility of the finding that face-to-face collaboration reduces creativity.

An implication of this finding is that future research will need to peer into the black box of group dynamics and structured analytic techniques. While drawing causal inferences about group behavior and social conformity was not an aim of this research, anecdotal data from the experiment backed this point as some study participants were more likely to be active, at times dominating the conversation while others remained quiet. Future research will have to delve into this question to tease out the relationship between structured analytic techniques, group collaboration, and the potential limitations on creative thinking. In the meantime, and against the grain of common wisdom, analysts should eschew face-to-face collaboration at the early, creative phases of a project. Instead, analysts should work in nominal groups, where ideas and judgments are generated individually and then combined by a group facilitator or mechanical aggregator, such as a software program (Armstrong 2006).

Another option is to conduct further research on alternative approaches to generate the fullest set of hypotheses possible. In the public policy literature, Dunn (2002) developed one is called “Boundary Analysis,” which was used in this study to examine experts judgments and determine hypothesis novelty. The first step in Boundary Analysis is the specification of the analytic problem. For example, “what are the likely outcomes of the Syrian Civil War?” Next, analysts sample data sources that hold hypotheses related to the analytic question, in this case different outcomes of the Syrian war. A common source of hypotheses can be found in open source documents, such as news reports or subject matter experts. As analysts collect hypotheses the set of unique hypotheses will initially grow exponentially with each document or expert interviewed then rapidly decrease and level-off. This rapid leveling-off is due to Bradford’s Law, an empirical regularity that states that after searching a few key sources, the analyst will have attained nearly all of unique hypotheses, an almost full set of hypotheses. As discussed above, the distribution of hypotheses almost perfectly conformed to Bradford’s Law: after the 19th expert nearly all hypotheses had been mentioned. The curved line in figure 6.13 represents the expected distribution of cumulative unique hypotheses based on Bradford’s law and the squiggly line the observed distribution. The fit is nearly perfect ($R^2 = .98$). After around 20 experts or sources the leveling off hypotheses lets the analyst know that there is a nearly full set of hypotheses.

Figure 6.11 Cumulative Frequency of Hypotheses



6.3.2 Beyond Confirmation Bias: Evidence Weighting

Wason's (1968) pioneering lab experiment found that study participants sought to confirm, rather than disconfirm hypotheses. After Heuer's landmark *Psychology of Intelligence Analysis* (1999), a fundamental problem, at least in the eyes of intelligence reformers, has been to mitigate confirmation bias, or in other words, to try to get analysts to try to disconfirm their favored hypotheses. The systematic review and the results of this experiment suggested the technique seems to at least modestly decrease focus on a single or small set of hypotheses. This finding was also apparent in the variation in the judgments concerning the percentage of weapons to be destroyed by the June 30th deadline. However, the experiment supports another finding from the systematic review: *ACH's ability to get analysts to disconfirm their hypotheses appears to have nothing to do with improving judgment accuracy.*

For some intelligence reformers and promoters of ACH, the lack of evidence for improvements in accuracy might not be dispiriting. Randy Pherson, a consultant and trainer in structured analytic techniques argues that using ACH to make intelligence estimates is a “bridge too far” for “98 percent” of analysts.⁵¹ While Pherson’s sentiment is understandable given the limitations of forming accurate judgments in intelligence analysis, improvements are possible. One way that Wheaton (2014) discusses improving analysis is training analysts to understand how to weight evidence. One such example that could be applied to techniques beyond ACH, is a source credibility scale for open source intelligence, developed at the National Cryptologic School (Norman 2001).⁵² The scale includes 12 items, including whether the source can corroborated by other sources and whether the author is reputable, among others. Improving this scale and expanding training efforts could be one way to improve the accuracy of intelligence judgments and advance the conversation beyond simply avoiding confirmation of a favored hypothesis.

6.3.3 Cognitive Reasoning Style and Structured Analytic Techniques

In light of the importance of cognitive reasoning styles for accuracy from other research (Tetlock 2005), the results for study hypothesis 4 were unexpected. There a couple reasons that might explain why cognitive reasoning style did not have an interaction an

⁵¹ Interview with Randy Pherson, Washington, D.C., June 2014.

⁵² Norman’s thesis as well as a source credibility scale are available at: <https://sites.google.com/site/daxrnorman2/analysis>

effect. First, the techniques used in this study are different in form and function than those used in Tetlock's study. In his study Tetlock utilized a version of Alternative Futures Analysis, an imaginative technique which different from the techniques used in this experiment. It is possible, that if this study used a similar imaginative technique, such as Brainstorming, the Tetlock's results might have been replicated. Still, a plausible rebuttal is that the hypothesis generating step in ACH should have mimicked Brainstorming, thus getting around this limitation.

Another possible explanation is the validity of the cognitive reasoning style questionnaire used to assess whether respondents were "foxes" or "hedgehogs" from Tetlock's *Expert Political Judgment* (2005). This magisterial work has been criticized on numerous grounds. In fact, Tetlock's research on political judgment has drawn so much comment that an entire issue of *Critical Review* (2010) was devoted to essays critiquing his findings. Absent from these critiques is an important but mundane element of his argument: how do we measure the difference between hedgehogness and foxiness? To use psychometric jargon, how do we know that the questionnaire instrument Tetlock used to assess reasoning styles was valid? Given the importance of the hedgehog-fox scale for his Tetlock's thesis, it is odd that he does not report validity or reliability statistics in the methodological appendix in the sections discussing the cognitive reasoning style questionnaire (p. 240 and p. 269).

To examine the validity of the cognitive reasoning style questionnaire, participants in this study also took the Myers-Briggs type indicator (MBTI) short-form. Then participants' scores were calculated for the judging and perceiving dimension, a

dimension of the MBTI that I hypothesized should closely correlate with the hedgehog/fox scale. Those closer to the judging end of the scale are individuals who prefer to “live in a planned, orderly way, seeking to regulate and manage their lives. [those closer to the judging end want] to make decisions, come to closure, and move on” (Myers 1998). In other words, the higher the participant scores in “judging” the more hedgehog-like he or she should be. Those closer to the perceiving end of the scale tend “to live in a flexible, spontaneous way, seeking to experience and understand life, rather than control it.” (Myers 1998). Again, the higher the perceiving score the more fox-like the participant should be. Consequently, it is expected that both scales would be correlated. However, a Pearson product-moment correlation coefficient was computed to assess the relationship between participant’s scores on the MBTI and hedgehog-fox questionnaire. There was no statistically significant correlation between the two variables ($r=.079$, $n=24$, $p=.713$). These results are surprising because there should be some level of convergence between the two measures, thus demonstrating convergent validity of the hedgehog-fox measure. Future research will need to explore the validity of foxness and hedgehogness with a larger number of participants.

6.4 SUMMARY AND CONCLUSIONS

The results of the study partially confirmed the finding that face-to-face collaboration limits idea quantity: the ACH group saw a decrease in the quantity of their ideas and the Indicators group saw no increase or decrease. These results cohere with results from the systematic review suggesting that face-to-face collaboration is unhelpful for encouraging creativity. To address problem, analysts should work in nominal or non-interacting groups.

Another study hypothesis was that there would be no relation between whether participants confirmed a favored hypothesis and forecasting accuracy. The study result supports findings from the ACH studies in the systematic review: ACH assisted participants to some extent in avoiding focus on a favored hypothesis, but it did not improve judgment accuracy. Results from the evaluation of Indicators were disappointing. The technique did not significantly improve the rigor or accuracy of analysis. The results of this study suggest there is no interaction effect between using structured analytic techniques as represented by ACH and Indicators, with cognitive reasoning style. All study participants, regardless of cognitive reasoning style, were equally affected by the two techniques. In particular, both those with open and closed reasoning styles saw the same narrowing effect on the quality and quantity of hypotheses. This result could be explained by the techniques used in this experiment that do not encourage analysts to think of many possible hypotheses (as Tetlock's

experiment required). Future research will be needed examine the construct validity of Tetlock's scale, specifically testing its reliability.

7.0 CHAPTER 7: EVIDENCE-BASED PRINCIPLES FOR IMPROVING INTELLIGENCE ANALYSIS

Research suggests that foreign affairs analysis is weak and even the best experts are highly fallible. In fact, the highest performing analysts are accurate less than 35 percent of the time (Tetlock 2005). This is problematic for statesmen and intelligence analysts alike, trying to make sense of a messy and chaotic world. Structured Analytic Techniques are one option for improving foreign affairs analysis. The imperative to use these methods was enshrined in the Intelligence Reform Act (2004) which mandates analysts use the techniques (§ 1017). This research sought to understand how the techniques have been implemented in the IC and make a modest attempt to evaluate a subset of 12 techniques listed in the U.S. Government's *Analytic Tradecraft Primer* (2009).

The research is based on semi-structured interviews conducted with 5 intelligence experts and a survey of 80 analysts at an IC agency along with follow-up interviews with 15 analysts. Approximately a third of analysts report never using the techniques on the job. The main factors related to the use of the techniques were training in and to a lesser extent, perception of the techniques. Although a range of professional opinion holds that time pressure is the main reason analysts do not use the techniques, there was not a

statistically significant relation between the time pressure under which analysts work and their use of the techniques.

Questions about effectiveness of the techniques were answered in part by employing a “systematic review,” a relatively new but well-accepted methodology of synthesizing a large body of research. After reviewing a sample of more than 2,000 studies of the use of the techniques, the evidence suggests that there is moderate to strong evidence affirming the efficacy of using three of the twelve core techniques (these techniques are “Analysis of Competing Hypotheses,” “Brainstorming,” and “Devil’s Advocacy”). There were three specific findings: face-to-face collaboration decreases creativity, evidence weighting appears to be more important for attaining accuracy than seeking disconfirming evidence, and conflict tends to improve the quality of analysis. Using a mixed-methods approach the research also employed an experiment with 21 graduate intelligence studies students, which supported the first two findings of the systematic review.

The evidence on the implementation and effectiveness of the techniques converge on a set of evidence-based principles for improving intelligence analysis covered in section 7.1. Each principle is backed up with multiple pieces of evidence from the survey, interviews, systematic review, and experiment. The description of each principle includes implications for the IC and opportunities for future research to improve and extend the knowledge base on “what works.” The chapter closes with a short summary of the main arguments.

7.1 EVIDENCE-BASED PRINCIPLES FOR INTELLIGENCE ANALYSIS

7.1.1 Training and the Value of Evidence

Principle 1: Implementing structured analytic techniques requires training analysts in their use and providing evidence that the techniques work. The survey and interviews conducted with INR analysts suggested that training is moderately correlated with the use of structured analytic techniques (chi square=13.593, $p=0.001$, Cramér's $V=0.412$). Another variable positively correlated with the use of the techniques, although to a lesser degree, is whether analysts perceive the techniques to be effective for improving the rigor of analysis (chi square=3.83 $p=.049$, Cramér's $V=0.225$). Other variables cited in the literature, such as age of the analysts (chi square= 0.009 $p=0.924$) and time pressure (chi square= 0.616, $p=0.735$) are not statistically significant relation with the use of the techniques. The finding on time pressure is notable because professional opinion holds that analysts are too busy to use the techniques (Heuer 1999, p. 85-86; Folker 2000; Khalsa 2009).

Combined, these results speak to a simple but important truth: the key to implementing the techniques is to provide quality training and a compelling reason to analysts that they should use the techniques. As the study suggested, training in the techniques is not uniform in the IC, and indeed, almost non-existent in INR. Implementing the techniques will require inter-agency collaboration and a willingness to explore and consider what constitutes “quality training” and follow through with

instruction that translates into practical skills. As the study of INR found, when an analyst considered training to be high quality, they were far more likely to use the techniques. Still, if the push for training cannot be made in the IC, the new degree granting intelligence programs may also play a role. Since the September 11th attacks more than 13 programs have been founded and some are experimenting with curriculum for teaching analytic methodologies, including structured analytic techniques (Crosston and Coulthart, n.d.).

State Department's Bureau of Intelligence and Research appears to differ from many other agencies because of its, size, culture, and performance. Therefore future research should determine the extent to which these variables affect the preference to use structured analytic techniques at other, larger agencies such as the Central Intelligence Agency and National Security Agency. Future research could examine state and local intelligence agencies as well to determine how often techniques are used and why. The survey methodology used in this research could easily be replicated in other agencies and follow-up interviews could be used to cross-validate the findings of the statistical analysis.

Another question for future research is to examine how the techniques are used. As the findings of this study suggested, training plays a significant role in whether the techniques are used, and presumably, how well they are used. An anecdotal report from some preliminary research done for this project suggested that at some IC agencies the techniques are sometimes used to please supervisors rather than to improve analysis. For example, one Defense Intelligence Agency analyst reported writing a report *then afterward*

using a technique to reach a predetermined conclusion.⁵³ Understanding how techniques are used (or abused) would probably require ethnographic fieldwork and participant observation. While access would be difficult to attain, it is entirely possible, as previous fieldwork in the IC has ably demonstrated, most notably by Johnston (2005).

7.1.2 Avoid Face-to-Face Collaboration

Principle 2: In the idea generation phase of a project, analysts should work individually, followed by a systematic method to aggregate ideas. Across all eight of the moderate and highly credible evaluative studies of Brainstorming, group collaboration limited the quantity and quality of ideas. This conclusion was also supported in the ACH experiment. The implication of this finding is that if analysts wish to use Brainstorming to produce more and better ideas, they should work independently and then pool their efforts using a designated, non-partial facilitator. Attempting to use a collaborative group to generate a pool of ideas, as suggested in the Tradecraft Primer (2009, pp. 27-29) is unlikely to produce the best outcome. More generally, this principle speaks to carefully applying collaboration, an area of growing emphasis in the IC (Medina 2008, p. 246).

An upshot of this finding is that intelligence agencies need to rethink how they conduct collaboration in the idea generation phase of projects. The survey of INR in [Chapter 4](#) suggested that most, if not all, Brainstorming is done through face-to-face

⁵³ Interview with DIA analyst, Washington, D.C., May 2013.

collaboration. Survey respondents also reported using a form of semi-structured Brainstorming in inter-agency discussion between the Central Intelligence Agency and National Intelligence Council.⁵⁴ Indeed, face-to-face collaboration is common throughout the IC for generating ideas and making sense of data (Johnston 2005, p. 5). Armstrong (2006) suggests that to improve forecasting accuracy, groups should use alternative means of combining ideas. He discusses the value of alternatives such as the Delphi method, a round-based communication system that was developed to address group dynamics and conformity that reduce creativity.⁵⁵ While there is some discussion in the intelligence literature of the Delphi method and government training materials for this technique exist, there are no detailed accounts of it being used in the IC.

Beyond these methods, the limits of face-to-face collaboration opens up research opportunities for introducing and developing new techniques. One such technique that could be adapted for intelligence is Boundary Analysis (Dunn 2001, 2012). As the name implies, Boundary Analysis is a technique to determine the analytic ‘boundaries’ of a problem, for example, in this case the number of plausible hypotheses. The first step in Boundary Analysis is the specification of the analytic problem. For example, “what are the likely outcomes of the Syrian Civil War?” Next, analysts sample data sources that set forth hypotheses related to the analytic question. A common source of hypotheses can be

⁵⁴ Interview with INR analysts, Washington, D.C., February 24-26, 2014.

⁵⁵ For an extended discussion of Delphi, see: Linstone, Harold A., and Murray Turoff, eds. *The Delphi method: Techniques and applications*. Vol. 29. Reading, MA: Addison-Wesley, 1975.

found in open source documents, such as news reports. Once the data are compiled, they can be mined by coding each unique hypothesis.

Future research could validate the technique and refine the procedures. For example, the technique has also been used in other intelligence contexts. Coulthart and Rickabaugh (2014) applied Boundary Analysis to understand if they could retrospectively identify a “black swan” hypothesis in a real-world criminal intelligence case. While the technique did not identify the black swan hypothesis, a full set of hypotheses was generated that almost perfectly conformed to Bradford’s law. Another interesting finding was that almost the entire distribution of hypotheses was mentioned within a two week period, while the case documents covered several months of speculation by journalists and bloggers. This finding, if replicable in other cases, is important because Boundary Analysis could assist analysts using open-source intelligence on emerging security threats to narrow and focus their analysis. Future research can refine the procedures of Boundary Analysis. In the aforementioned study, the entire extraction and enumeration process required a single coder approximately 8-10 hours. Adding additional coders could reduce this time and ensure coding reliability through calculating simple inter-coder reliability scores.

7.1.3 Focus on Evidence Weighting

Principle 3: To improve judgment accuracy, analysts must update their beliefs through carefully weighting evidence. The experiment in [Chapter 6](#), confirms the findings of Brasfield (2009)

and Wheaton (2014), who conclude that ACH assisted participants in avoiding focus on a favorite hypothesis, but it did not improve judgment accuracy. The studies on ACH from the systematic review suggested a new direction for research examining the role of evidence weighting. Wheaton (2014), describing the outcome of his study, concludes that “accurate estimates came from analysts who either a) intuitively weighted evidence without the help of a decision tool or b) were instructed how to use the decision tool with special focus on ... evidence weighting.” Wheaton concludes that perhaps the negative impact of bias may be less about searching for disconfirming evidence than about weighting such evidence, once discovered. The importance of weighting sources and evidence confirms Tetlock’s (2005, pp. 120-141) finding that the best forecasters are those who carefully updated their beliefs on the basis of new evidence.

The implication of these findings is that analysts will need to be trained to think beyond disconfirming their beliefs to include evidence weighting. Currently, the approach to teaching ACH and analytic methods generally, seems to be focused on seeking to disconfirm hypotheses. For example, the syllabus for a structured analytic techniques course offered by a large government contractor notes that it will explore “identification of alternative explanations and evaluation of all evidence that will *disconfirm* rather than confirm hypothesis” [emphasis added] (Lockheed Martin n.d.) Another syllabus from the National Security Agency states “people seek to confirm the first answer to a problem they discover, selectively using evidence to support that position” (Moore 2007). Undoubtedly, seeking disconfirming evidence is helpful because there appears to be a natural inclination to confirm prior beliefs (Wason 1968) and

disconfirmation has a long tradition in the philosophy of science through the work of Karl Popper (1953). The next step is to teach and aid analysts in thinking about how to assess the credibility and quality of sources of evidence.

One avenue for future research is to develop instruments or techniques to assist in evidence weighting. An example is an open source intelligence credibility scale developed at the National Cryptologic School (Norman 2001). The scale is out of date and provides no reliability measures so that one analyst may use the scale in a completely different way than another. For example, one item on the scale asks whether the author is “reputable,” a vague term that is open to wide interpretation. Future research should refine this scale and validate it among intelligence analysts. Additionally, future research can extend work underway by Tim van Gelder (2009) and others to implement argument mapping into the analytic process. Through applying structured forms of argumentation, such as the Toulmin argument model (1958), analysts can examine the quality and credibility of their analytic judgments.

7.1.4 Harness Conflict Carefully

Principle 4: Competitive analysis is helpful for improving the rigor and accuracy of analysis under certain conditions. Since the 1970s the IC has experimented with competitive analysis. The idea is that “estimative processes” can be sharpened when they are driven by the clash of competing ideas in a structured format.” (Mitchell 2006, p. 145). In other words, harnessing structured conflict should improve the analytic process and increase the

accuracy of judgments, an argument that appears to have its origins in the business and management literature (Churchman 1971; Mitroff and Turoff 1973; Mitroff and Emshoff 1979). Not surprisingly then, the systematic review found eight evaluative studies from the management and business literature that had at least moderately strong research designs for one competitive analysis technique: Devil's Advocacy. Across nearly all of the studies, conflict-based approaches (which includes Devil's Advocacy) outperformed consensus methods in improving judgment accuracy and validating assumptions.

The research from the systematic review suggests that there are circumstances where the use of conflict is most effective. Conflict inducing techniques appear to most improve analysis when experts' assumptions about the analytic task are invalid. In order to manipulate assumptions in these studies, researchers provided control and treatment groups using consensus and conflict-based approaches with correct and incorrect assumptions about the analytic task. According to a meta-analysis by Schwenk (1984, p. 306), "when the assumptions of the [task] are not correct, the conflict introduced by [conflict approaches] improves decision-making." Given the vagaries surrounding assumptions in intelligence analysis, conflict approaches such as Devil's Advocacy could be helpful. Task type also mattered for the effectiveness of Devil's Advocacy: when group performance is determined by the aggregation of individual effort of all group members, the technique may hamper the analytical process (Murrell *et al.* 1993). Interestingly, while conflict seems to improve analysis, Devil's Advocacy can lead participants to view the analytic process as less satisfying than consensus-based methods, thus possibly leading analysts to avoid using the technique (Tung 1992).

The implication of research is that competitive analysis holds great promise for intelligence analysis, but efforts must be made to carefully implement techniques. One anecdote highlights the problem: the infamous Team A/Team B on the Soviet Union's military strength and the use of competitive analysis in the lead-up to the Iraq War. According to Mitchell (2006, p. 144) the Iraq exercise was used to subvert "the competitive intelligence analysis process, where unofficial intelligence boutiques 'stovepiped' misleading intelligence assessments directly to policy-makers and undercut intelligence community input that ran counter to the White House's preconceived preventive war of choice against Iraq." Additionally, there are the very real problems of implementing conflict-driven techniques, such as Devil's Advocacy. One Central Intelligence Agency analytic methodologist interviewed for this research stated he avoided the technique because he felt the larger group would "route" around the devil's advocate.⁵⁶ Future research should examine how competitive analysis can be fruitfully implemented and determinant when the technique has been effective, or counter-productive. Similar to the above discussion on understanding how techniques are used in the IC, future research could use ethnographic and participant observation methods to understand potential barriers in the analytic process.

⁵⁶ Interview with a CIA methodologist, June, 2014.

7.2 SEEKING TO BE “APPROXIMATELY RIGHT”: SUMMARY AND POLICY IMPLICATIONS

The eminent statistician John Tukey wrote he would rather be "be approximately right rather than exactly wrong." This dissertation attempted to explore the limitations and potential correctives of foreign affairs analysis, an area of human endeavor where we are often “exactly wrong.” Understanding the extent to which structured analytic techniques might improve analysis required not only looking at effectiveness but also the application of the techniques in a secretive environment. The result was in-depth interviews with five intelligence experts and a survey potentially generalizable to an IC agency – the first of its kind publicly available. Effectiveness was assessed with the combination of a systematic review sampled from more than 2,000 research studies and a field experiment involving 21 graduate intelligence studies students.

The findings from this dissertation provided four evidence-based principles. First, implementing structured analytic techniques requires training analysts in their use and providing evidence that the techniques work. Second, during the idea generation phase of a project, analysts should work individually, followed by a systematic method to aggregate ideas. Third, to improve judgment accuracy, analysts must update their beliefs through carefully weighting evidence. Lastly, competitive analysis is helpful for improving the rigor and accuracy of analysis but under certain conditions and with task types.

Combined, the findings and principles lead to a couple recommendations for how the IC could increase the probability of being “approximately right”:

- *Fund IC centers for analytic excellence.*

There are significant gaps in our understanding how to reliably produce quality intelligence analysis. To address this problem, quality research on “what works” needs to be increased. The Intelligence Advanced Research Projects Activity (IARPA) has done an excellent job stimulating innovative research but their project-based approach to funding research should be supplemented by more long-term research initiatives. A next step would be to develop centers for analytic excellence at select universities. Creation of these centers is not unprecedented as the IC already funds Intelligence Centers for Academic Excellence program (IC-CAE), which are used to increase the diversity of human capital available to the IC (Defense Intelligence Agency, 2011). The analytic centers for academic excellence could either be added on to existing IC-CAEs or created as standalone entities.

- *Incorporate evidence into the training process.*

Knowledge generated by evidence-based intelligence analysis needs to be implemented into training as it has been done in a variety of fields ranging from medicine to law enforcement. As the findings above suggest, training and perceptions of the techniques are important for influencing how analysts do their jobs. Perhaps just as important, bringing insights from evaluative research into the IC’s training curriculum

could help analysts generate higher quality analysis. For example, the implication of the principle on competitive analysis is that analysts should be aware of when to use techniques that harness conflict. The results of one study suggest, in tasks where each group member contributes information to create one combined judgment, using a technique like Devil's Advocacy can harm the quality of analysis (Murrell *et al.* 1993). Integrating evidence from evaluative studies can be accomplished by redeveloping and updating the curriculum at intelligence training centers. One way to bring evidence into the training curriculum is for the IC to rely on the proposed centers for analytic excellence or to use the existing IC-CAE programs to develop curriculum.

These two recommendations are easily implemented and as Tetlock and Mellers (2011, p. 543) note, "The IC does not have to lower the probability of multibillion-dollar fiascos by much to recoup a multimillion-dollar investment." In other words, meagre investments in improving analysis can have a large impact. Still, funding is one matter and improving analysis is another. Future research and reform of the intelligence analysis process will need to find ways to push the frontiers further to improve analysis of foreign affairs.

8.0 METHODOLOGICAL APPENDIX A: SURVEY AND INTERVIEW PROCEDURES

Intelligence Expert Interviews

5 intelligence expert informants were identified using a snowball sampling strategy. 5 respondents did not respond to inquiries.

Each was interviewed with the following protocol:

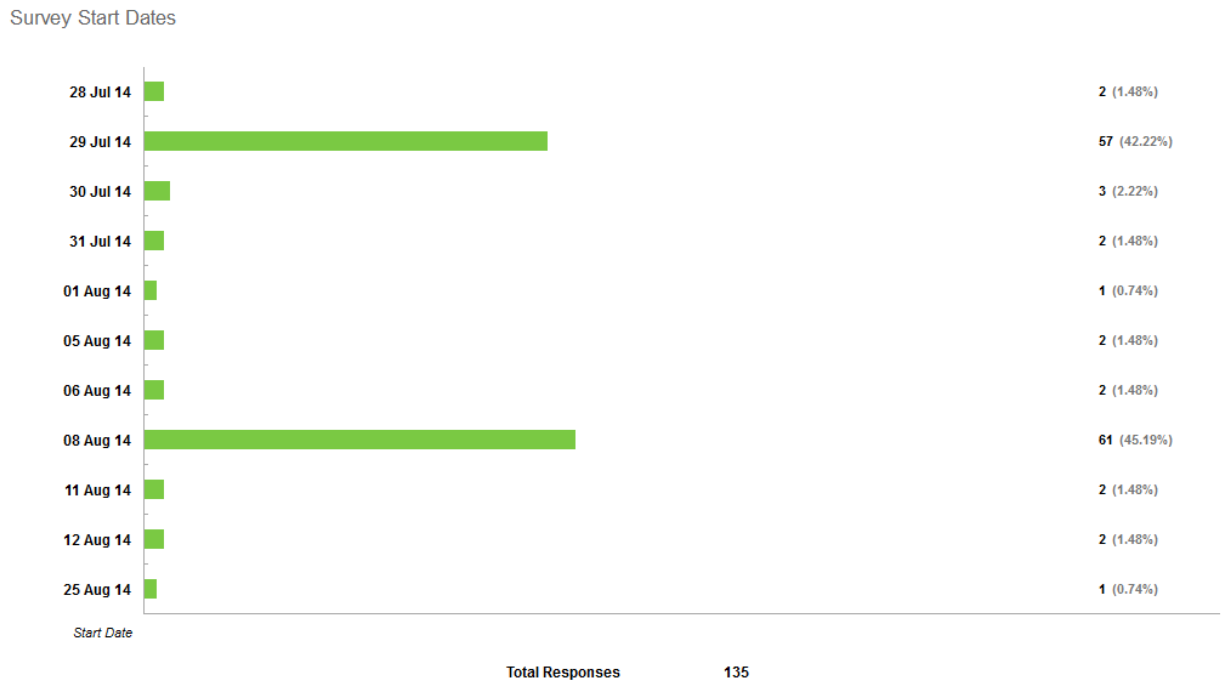
Confidentiality Statement: My name is Steve Coulthart and I am researching how structured analytic techniques are utilized in the intelligence community, such as alternative competing hypotheses, red teaming, and alternative futures analysis, among others. This interview will require roughly 20-30minutes and address your background and expertise with these techniques. Be assured that your responses and those of others are entirely confidential. Your participation is voluntary and you may choose to end the interview at any time.

Semi-Structured Questions:

- 1) Could you start by telling me about your background in intelligence analysis?
- 2) Since the 2004 Intelligence Reform act, have you seen any changes in the analytic culture of the IC?
- 3) How often do you think analysts are utilizing structured analytic techniques on the job?
- 4) What factors do you think affect the use of the techniques?
- 5) Do you think there is variation in the use of the techniques in the IC?
- 6) How well prepared do you believe analysts are for using structured analytic techniques?

Bureau of Intelligence and Research (INR) Survey

An email was sent out to a list-serv at INR in two rounds: July 28th, 2014 and August 8th.



137 surveys were initiated, including 95 with at least 80 percent or more of the questions completed. Of the 95 responses, 80 respondents could be described as analysts under their current or former job titles.

The survey protocol contained the following information and questions:

Confidentiality script: The purpose of this survey is to explore your views as an intelligence analyst on analytic tradecraft practices. I will also be surveying other analysts, asking them to complete this brief 3 minute online questionnaire. The survey asks questions about your background as an analyst and your views on such analytic tradecraft practices, such as structured analytic techniques. Please leave your email address if you would be willing to participate in a follow-on interview.

There are no foreseeable risks associated with this survey. I will be pleased to send you a summary of the survey results, but be assured that your responses and those of others are entirely confidential. Your participation is voluntary and you may choose not

to participate. The survey is being conducted by Steven Coulthart, a doctoral candidate at the University of Pittsburgh. If you have any questions I can be contacted at (315) 264-0917 or SJC62@pitt.edu.

Please leave your email address if you would be willing to participate in a follow-on interview.

Email (optional):

1) How many years have you been an analyst?

-[open-ended]

2) What agency do you work for?

-[open-ended]

3) What is your current job title?

-[open-ended]

4) On average, how long does it take you to complete an analytic project?

-A day or two

-A week or two

-A month or two

-Three months or more

5) Several kinds of “structured analytic techniques” are used by analysts. These include Analysis of Competing Hypotheses (ACH), link analysis, and red teaming, among others. Have you received any training in any of these structured analytic techniques?

-Yes

-No

6) How well would you say the training prepared you to apply these techniques?

-Very well

-Somewhat well

-Not very well

-Not well at all

7) How often do you use structured analytic techniques on the job?

-Occasionally

-Rarely

-Never

8) To what extent, on average, do you think structured analytic techniques help analysts think in a more effective way (e.g. consider new perspectives, challenge mental models, etc) ?

- A great deal
- A fair amount
- A little
- Not at all

9) To what extent, on average, do you think structured analytic techniques help analysts be more accurate or "right" in their analytic judgments?

- A great deal
- A fair amount
- A little
- Not at all

10) It has been 10 years since major analytic reforms have been put in place (for example, under the 2004 Intelligence Reform Act). What do you believe the impact of these reforms has been?

- [open ended]

Bureau of Intelligence and Research Survey Follow-up Interviews

Confidentiality script: My name is Steve Coulthart and I am researching how structured analytic techniques are utilized in the intelligence community, such as alternative competing hypotheses, red teaming, and alternative futures analysis, among others. This interview will require roughly 15 minutes and address your background and expertise with these techniques. Be assured that your responses and those of others are entirely confidential. Your participation is voluntary and you may choose to end the interview at any time.

1) How well prepared do you believe analysts are for using structured analytic techniques? Have you received training in structured analytic techniques?

2) In your opinion, what is the perceived value of structured analytic techniques among the "average" analyst at INR?

3) What factors do you think influence use of structured analytic techniques?

4) The intelligence community appears to be seeing a demographic shift, with an influx of younger analysts. Do you think this will have an impact on the use of structured analytic techniques? Why or why not?

Analysis

Data were analyzed using non-parametric statistics including chi square tests and measures of association for categorical data. For each of the tests between variables categories were collapsed to ensure the cell counts did not drop below the required 5.

Qualitative data were analyzed with qualitative data analysis software to generate a set of codes:

| Nodes | | | |
|-------|---|---------|------------|
| | Name | Sources | References |
| [-] | Analytic Task | 0 | 0 |
| | Problem-Analytic Task Type | 5 | 8 |
| | Resources needed to use Structured Analytic T | 4 | 4 |
| | Time Pressure | 3 | 3 |
| [-] | Individual-Level Factors | 0 | 0 |
| | A Graying IC | 7 | 8 |
| | Generalists versus Experts | 4 | 4 |
| | Perceptions of Effectiveness | 7 | 10 |
| [-] | Institutional Factors | 0 | 0 |
| | Generalists-Expertise Culture | 4 | 10 |
| | Institutional Pressure | 1 | 1 |
| | Institutional Size | 4 | 8 |
| | Institutional Support | 2 | 3 |
| | Interagency Cross Polination | 8 | 14 |
| [-] | Training | 3 | 7 |
| | Informal Training | 2 | 2 |
| | Lack of Training | 3 | 4 |
| | Use of Structured Analytic Techniques | 4 | 5 |

PROCEDURES

Search Keywords and Settings

- Searched “scenario planning” in lieu of alternative futures analysis, a less commonly used term for the technique. Search parameters: did not include citations and search terms had to occur in the title in order to narrow the results to most applicable studies. Only English studies were included. Yielded 753 studies
- Searched "red teaming" in the title, exclude the terms “automated” “computational” and other search terms to exclude studies focused on non-analytic red teaming. Only English studies were included. Did not include citations. Yielded 21 research studies
- Searched "Team B" in the title , exclude the terms “automated” “computational” and other search terms to exclude studies focused on non-analytic Team B. Only English studies were included. Yielded 17 research studies
- Search "analysis of competing hypotheses" in the title. Search parameter: do not include citations. Only English studies were included. Yielded 20 research studies.
- Search “brainstorming.” Search parameters: did not include citations, patents, and search terms had to occur in the title in order to narrow the results to most applicable studies. Only English studies were included. Yielded 838 studies.
- Other techniques did not yield results: Key Assumptions Check, Quality of Information Check, High-Impact/Low-Probability, Indicators of Signpost/Change, “What If?” Analysis, and Outside-In Thinking

Randomized Stratified Sampling Procedure (scenario planning is used as an example):

STEP ONE: Define the population: all studies returned by the search terms (see STEP THREE)

STEP TWO: Choose the relevant stratification: Stratification is based the assumption that the most cited studies will be listed in the top 20% (“highly relevant articles”) and less relevant in the remaining 80%. See Beel and Gipp 2009 for an empirical test of this claim.

STEP THREE: List the population: the list of the population is available only with the entering of the Google search (note this list may change daily). The population is 753 studies generated by the search terms “scenario planning.” Search parameters: did not include citations and search terms had to occur in the title in order to narrow the results to most applicable studies. Only English studies were included.

STEP FOUR: List the population according to the chosen stratification: The first 8 pages of results (20 studies per page) is 20% of the population. The remaining 80% contains the rest of the results.

STEP FIVE: The sample size was set by using a calculator: assuming a population of 797 (the number of returned results), a 95% confidence level and a margin of error of 10%, the sample is 86

STEP SIX: Calculate a proportionate stratification:
Will need 17 highly relevant articles (from the top 20%)
Will need 69 less relevant articles (from the bottom 80%)

STEP SEVEN: Sample and download articles: Using a random number generator to select two sets of numbers page number and study position. For example, if there are pages possible and there are 20 studies per page, the randomly generated number 220 would require downloading the study from the 20th position on page 2.
Following this process 86 articles were downloaded

Excluded Studies and Reasons

Total studies excluded: 213 (46 included)

Total time required: 10 hours

To be included the study must have ALL of the following criteria:

- 1) The author provided some report, verbal or numerical of the technique on some type of outcome.
- 2) Made some attempt to evaluate a specific, identifiable instance of the technique. As a result, hypothetical reports (e.g. “ACH could be helpful to intelligence analysts...”)
were excluded

In addition to the above criteria, studies were also excluded on the following traits:

- Was a book – only articles and shorter manuscripts were considered for logistical reasons (2 studies)
- Describes how to use the technique in an application but provides no evaluation of effectiveness (95 studies)
- Discusses issues related to the use of the technique (36 studies)
- Duplicates a previous observation (e.g. multiple studies covered the IC Team A/Team B exercise from the 1970s). In these cases, the study with the stronger design (operationalized by the study's MSMS score) was selected, if designs were identical the first sampled study was included. (4 studies)
- Excluded studies looking at educational outcomes, such as creative writing skills. While these outcomes are of partial interest, they are not closely related to outcomes related to intelligence analysis (6 studies)
- Proposes an improvement or modification of the technique-but no evaluation (13 studies)
- In another language (not available in English) (1 study)
- Was the comparison of one variation of a technique against another (28 studies compared one variation of a technique against another)
- Technique addressed in study was not one of those included in this study as defined in the *Tradecraft Primer* (27 studies)

List of excluded studies and criteria:

| Title | Authors | Technique | Year | Exclusion Reason |
|---|-----------------------|------------------------------|------|--|
| The Banach-Barkelw Brainstorming Book. | WJ Banach, AH Barkelw | Brainstorming | 1976 | Book |
| Brainstorming: Techniques for New Ideas | TR Cory | Brainstorming | 2003 | Book |
| From Tradecraft to Profession | Aldric Ludescher | ACH | 2010 | Describes how to use the technique in a particular application (coporate security\) |
| Estimating rural Pashtun settlement population in Arghandab district, Zabul | Kevin Stofan | ACH | 2009 | Describes how to use the technique in a particular application (geospatial intelligence) |
| Envisioning the future of | E Oteros-Rozas, B | Alternative Futures Analysis | 2013 | Describes how to use the technique in a particular |

| | | | | |
|--|--|------------------------------|------|--|
| transhumant pastoralism through participatory scenario planning: a case study in Spain | Martín-López, CA López | | | application but provides no evaluation of effectiveness |
| Extension Facilitated Scenario Planning to Direct a Preferred Agriculture Future | JE Rowntree, MR Raven, JP Schwehofer, DD Buskirk | Alternative Futures Analysis | 2012 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness |
| Searching for the Essence of Red Teaming: Linearity Overcoming Rationality Toward Sensemaking | JW Bell | Red Teaming | 2010 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness |
| SCENARIO-BASED CONTRACTS: USING SCENARIO PLANNING AND VALUATION METHODS TO MANAGE RISK IN ALTERNATIVE DELIVERY METHODS | AA Amekudzi | Alternative Futures Analysis | 2005 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness |
| Scenario planning for simulation-based probabilistic risk assessment | Hamed Nejad | Alternative Futures Analysis | 2005 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness |
| Utilising scenario planning in the transport industry | JH Nell | Alternative Futures Analysis | 2003 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness |
| Red Gaming in Support of the War on | JH Moore, JB Whitley, RL Craff | Red Teaming | 2004 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness |
| Red Teaming: A Means to Military Transformation | John Sandoz | Red Teaming | 2001 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness |
| Red dawn: The emergence of a Red Teaming capability in The Canadian Forces | M Lauder | Red Teaming | 2009 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness |
| Scenario Planning as a Management Tool for Sustainable Aquaculture | W Jimma, D Adjei-Boateng, N Van Vosselen | Alternative Futures Analysis | nd | Describes how to use the technique in a particular application but provides no |

| | | | | |
|--|---|------------------------------|------|--|
| | | | | evaluation of effectiveness (Aquaculture) |
| Scenario Planning: The Future of Bosch Projects as Seen Through the Sugar Industry Lens | MA Madiba | Alternative Futures Analysis | 2010 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (Bosch projects) |
| Remembered Futures | Adam Cowart | Alternative Futures Analysis | 2011 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (business planning) |
| Dealing with the uncertainties of environmental change by adding | Phillip Walsh | Alternative Futures Analysis | 2005 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (business planning) |
| Scenario planning for climate change adaptation | SS Moore, NE Seavy, M Gerhart | Alternative Futures Analysis | 2013 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (climate change) |
| What if? Future Seas Scenario Planning and the Establishment of a Marine Reserve Network | Hammish Rennie | Alternative Futures Analysis | 2009 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (climate change) |
| Scenario planning for climate strategies development by integrating group Delphi, AHP and dynamic fuzzy cognitive maps | RBS Dolinšek | Alternative Futures Analysis | 2010 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (climate change) |
| Land Use-Transportation Scenario Planning in an Era of Global Climate Change | K Bartholomew, R Ewing | Alternative Futures Analysis | 2008 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (climate change) |
| SCENARIO PLANNING AS PART OF A REGIONAL CONSTRUCTION INNOVATION STRATEGY | C Abbott, P Chan | Alternative Futures Analysis | nd | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (construction) |
| Scenario planning for construction companies | R Soetanto, CI Goodier, SA Austin, ARJ Dainty | Alternative Futures Analysis | 2007 | Describes how to use the technique in a particular application but provides no |

| | | | | |
|--|--|------------------------------|------|--|
| | | | | evaluation of effectiveness (construction) |
| Using scenarios to develop crisis managers | Jason Moats, Thomas Chermack, Larry Dooley | Alternative Futures Analysis | 2008 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (disaster planning) |
| Planning and Executing Scenario Based Simulation Exercises: | D Van Niekerk, C Coetzee, D Botha | Alternative Futures Analysis | 2014 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (disaster planning) |
| Scenario modelling as a tool for planning sustainable urban energy systems | S Ben Amer | Alternative Futures Analysis | 2014 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (energy) |
| Imagining catastrophe: Scenario planning and the striving for epistemic security | U Tellmann | Alternative Futures Analysis | 2009 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (finance) |
| Scenario Planning for Building Coastal Resilience in the Face of Sea Level Rise | K McNamee, E Wisheropp, C Weinstein | Alternative Futures Analysis | 2014 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (flooding) |
| Portfolio design for investment companies | Payam Hanafizadeh, Abolfazl Kazazi and Azam Jalili Bolhasani | Alternative Futures Analysis | 2011 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (forecasting for Iran business) |
| Meeting the China challenge: Some insights from scenario-based planning | Richard Weitz | Alternative Futures Analysis | 2001 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (geopolitics) |
| Scenario Planning of Handheld Device | Y Lin | Alternative Futures Analysis | 2008 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (handheld devices) |
| Team 7: Applying automated red teaming in a maritime scenario | K Lin | Red Teaming | nd | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (harbor defense) |
| Save the Children UK: Southern Africa | Alan Whiteside, Su Erskin | Alternative Futures Analysis | 2002 | Describes how to use the technique in a particular application but provides no |

| | | | | |
|--|---|--|------|--|
| scenario planning paper | | | | evaluation of effectiveness (HIV/disease) |
| The President has no clothes: the case for broader application of Red Teaming within Homeland Security | AB Nettles | Red Teaming | 2010 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (homeland security) |
| Mapping of future Medical Universities program: scenario planning approach | A Allami, RQ Barqi | Alternative Futures Analysis | nd | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (hospitals) |
| Exploring Scenario Planning Processes | Fm Pastor | Alternative Futures Analysis | 2009 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (in 3 companies) |
| Scenario planning for the future of reference | Sarah Barbara Watstein | Alternative Futures Analysis | 2003 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (information sciences, reference) |
| Modeling a Policy for Managing Polio Vaccine in Japan: Scenario Planning based on System Dynamics | , K Hiyosh | Alternative Futures Analysis | nd | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (Japanese Polio situation) |
| Black swans to grey swans: revealing the uncertainty | AJ Masys | Red Teaming/Alternative Futures Analysis | 2012 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (law enforcement) |
| An Initial Conceptualization of Virtual Scenario Planning | Rochell McWhorter and Susan Lynham | Alternative Futures Analysis | 2014 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (leadership) |
| Scenario planning for libraries | Stuart Hannabuss | Alternative Futures Analysis | 2001 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (libraries) |
| Scenario planning for a library future | Steve O'Conner, Leonie Blair, Brenda Mconchie | Alternative Futures Analysis | 1997 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (libraries) |

| | | | | |
|---|---|------------------------------|------|--|
| Exploring the Future of Digital Reference through Scenario Planning | S Nicholson | Alternative Futures Analysis | 2003 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (library science) |
| Scenario Planning and Evaluation of Pricing Strategies in the Portuguese Bulk LPG Market | C Capelo, JF Dias | Alternative Futures Analysis | 2004 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (liquid petrol gas) |
| Analysis of Competing Hypothesis for Investigating Lone Wolf Terrorists | Lisa Kaati, Pontus Svenson | ACH | 2011 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (Lone Wolf terrorism) |
| Planning Future Technology Strategies Using Patent Information Analysis and Scenario Planning | JH Yoon, SC Cho | Alternative Futures Analysis | 2012 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (patents) |
| Managing Gas Assets through Simulation and Scenario Planning | Vannan, Mani, and Will Glass-Husain | Alternative Futures Analysis | 1999 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (petrol industry) |
| Using scenario planning in public health: anticipating alternative futures | JA Neiner, EH Howze, ML Greaney | Alternative Futures Analysis | 2004 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (public health) |
| Scenario analysis and strategic planning: practical applications for radiology practice | Frank James Lexa, Stephen Chan | Alternative Futures Analysis | 2012 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (radiology) |
| A Scenario Planning Report for Kinnexus | Federico de Silva Leon, Max Dunn, Hans Eberle | Alternative Futures Analysis | 2010 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (senior issues) |
| INTEGRATION OF SCENARIO PLANNING AND DECISION TREE ANALYSIS FOR NEW PRODUCT DEVELOPMENT: A CASE STUDY OF A SMARTPHONE ... | JZ Wu, KS Lina, CY Wub | Alternative Futures Analysis | 2015 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (smart phones) |

| | | | | |
|---|-------------------------------------|------------------------------|------|--|
| Scenario planning 2020 for Southern African economic empowerment | M Siwale | Alternative Futures Analysis | 2007 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (South African economy) |
| which ways can scenario planning contribute to the management of Dutch professional football clubs in case of relegation | Niels Wigbold | Alternative Futures Analysis | 2013 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (sports) |
| An Evaluation of Scenario Planning for Supply Chain Design | Yishai Boasson | Alternative Futures Analysis | 2000 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (supply chains) |
| Scenario Planning after Digital Switchover in S. Korea: The Use of Futures Wheel | JS Oh | Alternative Futures Analysis | 2009 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (technology) |
| Post Mission Red Teaming the Details of the Devil | C Matherly | Red Teaming | 2013 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (terrorism) |
| Red Teaming the Red Team | G Akins | Red Teaming | 2013 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (terrorism) |
| Using scenario planning to identify potential impacts of socio-demographic change on aspects of domestic tourism demand in Queensland in 2021 | PS Glover | Alternative Futures Analysis | 2006 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (tourism) |
| Scenario-based multiple criteria analysis for infrastructure policy impacts and planning | Matthew Schroeder and James Lambert | Alternative Futures Analysis | 2011 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (transportation) |
| Whether malevolent or negative, creativity is relevant to terrorism prevention: Lessons | Keith James and Damon Drown | Red Teaming | 2008 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness (trucking and terrorism)- |

| | | | | |
|--|---------------------------------|------------------------------|------|--|
| from 9/11 and hazardous material trucking | | | | provides pilot study but no report on effectiveness |
| Decision making under extreme uncertainty: blending quantitative modeling and scenario planning | Peter Kennedy and Robert Avilia | Alternative Futures Analysis | 2013 | Describes how to use the technique in a particular application but provides no evaluation of effectiveness-it provides lesson learned but no clear evaluation or report of the IMPACT of the technique |
| Detecting Deception by Analysis of Competing Hypotheses | C Elsaesser, FJ Stech | ACH | 2003 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Measuring Evidence During Criminal Defense Investigations Through Analysis of Competing Hypotheses (ACH) | JL Pennington | ACH | 2012 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| A Multinomial-Dirichlet Model for Analysis of Competing Hypotheses | KA Duncan, JL Wilson | ACH | 2008 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Midway Revisited: Detecting Deception by Analysis of Competing Hypotheses | FJ Stech, C Elsaesser | ACH | 2007 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| The missing link of crime analysis: a systematic approach to testing competing hypotheses | M Townsley, M Mann, K Garrett | ACH | 2011 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Applications of Analysis of Competing Hypotheses on Frontier Defense Intelligence Analysis | XIN Yongtao | ACH | 2008 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Developing the Next Generation of Scenario Planning Software | E Mueller, P lanning Softw | Alternative Futures Analysis | nd | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Exclusion Brainstorming | QR Code | Brainstorming | nd | Describes how to use the technique in an application but provides no evaluation of effectiveness |

| | | | | |
|--|--|---------------|------|--|
| Brainstorming | J Smolík, P Papiežová Vejvodová | Brainstorming | 2013 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Brainstorming in the College Classroom | WF Wetzler | Brainstorming | 1962 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Effect of Brainstorming Strategy on development the critical reading skills of French Language | W Al-Adl | Brainstorming | 2014 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Webstorming: Brainstorming in the Web | L Tarouco, S Amoretti, R Keller, S Garrido | Brainstorming | 2003 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Brainstorming as a prewriting activity | V Pfothenhauer | Brainstorming | 1982 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Reaching consensus through electronic brainstorming. | KA Scordo | Brainstorming | 1996 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| WordPlay: A table-top interface for collaborative brainstorming and decision making | S Hunter, P Maes | Brainstorming | 2008 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Proceedings of Brainstorming Session of Brainstorming Session of Brainstorming Session | A Pramanik | Brainstorming | 2011 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Active Brainstorming:-A systemic and systematic approach for idea generation | JE Kasser | Brainstorming | 2009 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Spiking Retina and Brainstorming | M Ebner | Brainstorming | nd | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Integrating TRIZ with value engineering: | DW Clarke | Brainstorming | 1999 | Describes how to use the technique in an application but |

| | | | | |
|---|----------------------------------|-----------------|------|--|
| discovering alternatives to traditional brainstorming and the selection and use of ideas | | | | provides no evaluation of effectiveness |
| Active Brainstorming:-A systemic and systematic approach for idea generation | JE Kasser | Brainstorming | 2009 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Enhancing English Learning through Brainstorming | H Houston | Brainstorming | nd | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Devil's Advocacy and the board: A modest proposal | CR Schwenk | Devils Advocacy | 1989 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Decision making models, Devil's Advocacy, and the control of corporate crime | MB Metzger, CR Schwenk | Devils Advocacy | 1990 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Devils Advocacy | Tw Roby | Devils Advocacy | 1998 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Devil's Advocacy and Dialectical Inquiry: Antidotes to Groupthink | F Lunenburg | Devils Advocacy | 2012 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Devil's Advocacy and patient choice | S Gallivan, M Utley | Devils Advocacy | 2004 | Describes how to use the technique in an application but provides no evaluation of effectiveness |
| Combining MMOWGLI Social Media Brainstorming with Lexical Link Analysis | Y Zhao, D Brutzman, DJ MacKinnon | Brainstorming | 2013 | Describes how to use the technique in an application but provides no evaluation of effectiveness (general) |
| Reading, thinking, brainstorming, writing and planning took place separately. Then sharing, | | Brainstorming | 1991 | Describes how to use the technique in an application but provides no evaluation of effectiveness (general) |
| he Role and Status of DoD Red Teaming Activiites | T Gold, B Hermann | Red Teaming | 2003 | Describes how to use the technique in an application but provides no evaluation of effectiveness (general) |

| | | | | |
|--|---------------------------------------|------------------------------|------|--|
| Red Teaming the Terrorist Threat | J Sinai | Red Teaming | 2003 | Describes how to use the technique in an application but provides no evaluation of effectiveness (general) |
| The use of red teaming in the corporate environment: A study of security management, vulnerabilities and defence | G Lane, D Brooks | Red Teaming | 2008 | Describes how to use the technique in an application but provides no evaluation of effectiveness (mining) |
| Three decision-making aids: brainstorming, nominal group, and Delphi technique. | AR McMurray | Brainstorming | 1994 | Describes how to use the technique in an application but provides no evaluation of effectiveness (nursing) |
| Does quantity generate quality? Testing the fundamental principle of brainstorming | AM Adánez | Brainstorming | 2005 | Describes how to use the technique in an application but provides no evaluation of effectiveness (nursing) |
| Learning from the future through scenario planning | Michael J Blyth | Alternative Futures Analysis | 2005 | Describes how to use the technique in general applications |
| Scenario Planning: Investing in Consciousness | D Pfenninger | Alternative Futures Analysis | 1998 | Describes how to use the technique in general applications but provides no evaluation of effectiveness |
| Directions in scenario planning literature–A review of the past decades | Varum, Celeste Amorim, and Carla Melo | Alternative Futures Analysis | 2010 | Describes how to use the technique in general applications but provides no evaluation of effectiveness |
| Decision Driven Scenario Planning for Process-Level Interventions | TJ Chemarck, TD Payne | Alternative Futures Analysis | nd | Discusses issues related to the use of the technique |
| Scenario Planning: An Innovative Approach to Strategy Development | Maree Conway | Alternative Futures Analysis | 2004 | Discusses issues related to the use of the technique |
| The Use and Value of Scenario Planning | JF Cardoso, MR Emes | Alternative Futures Analysis | 2014 | Discusses issues related to the use of the technique |
| Tailoring scenario planning to the company culture. | David Mason | Alternative Futures Analysis | 2003 | Discusses issues related to the use of the technique |
| Plausibility and Probabilty in Scenario Planning | Rafael Ramierz and Cynthia Selin | Alternative Futures Analysis | 2014 | Discusses issues related to the use of the technique |

| | | | | |
|---|-----------------------------|------------------------------|------|--|
| Mapping Public and Private Scenario Planning | Jay Ogilvy and Erik Smith | Alternative Futures Analysis | 2004 | Discusses issues related to the use of the technique |
| Exploring client scenarios associated with scenario planning | O Freeman, HM Pattinson | Alternative Futures Analysis | 2010 | Discusses issues related to the use of the technique |
| Decision making and planning under low levels of predictability: | George Wright, Paul Goodwin | Alternative Futures Analysis | 2009 | Discusses issues related to the use of the technique |
| Designing for the subconscious | RK Minas | Brainstorming | 2014 | Discusses issues related to the use of the technique |
| A Physical Artifact for Moderating and Analyzing Brainstorming Sessions | C Kakoulli | Brainstorming | 2011 | Discusses issues related to the use of the technique |
| Brainstorming Does Work for Agile | G Wilkstrand | Brainstorming | 2012 | Discusses issues related to the use of the technique |
| Quit Brainstorming and start Q-Storming® | J Dager | Brainstorming | 2011 | Discusses issues related to the use of the technique |
| Brainstorming Is Not Very Creative | J Baumgartner | Brainstorming | nd | Discusses issues related to the use of the technique |
| "But That Actually Happened!" Exploring the Speech Genre of Brainstorming | J Herbert | Brainstorming | 2013 | Discusses issues related to the use of the technique |
| Conducting Creativity Brainstorming Sessions in Small and Medium-Sized | US Murthy | Brainstorming | nd | Discusses issues related to the use of the technique |
| Making group brainstorming more effective | VR Brown, PB Paulus | Brainstorming | 2002 | Discusses issues related to the use of the technique |
| Brainstorming revisited: a question of context | T Rickards | Brainstorming | 1999 | Discusses issues related to the use of the technique |
| On creativity: A brainstorming session | U Bröckling | Brainstorming | 2006 | Discusses issues related to the use of the technique |
| Beta versus VHS and the acceptance of electronic brainstorming technology | AR Dennis, BA Reinicke | Brainstorming | 2004 | Discusses issues related to the use of the technique |
| Conducting Creativity | US Murthy | Brainstorming | 2009 | Discusses issues related to the use of the technique |

| | | | | |
|--|--|-----------------|------|--|
| Brainstorming Sessions in Small and Medium-Sized Enterprises | | | | |
| Brainstorming: An Act of Creativity in Vocational and Technical Education Curriculum in Nigerian Secondary Schools | SO Gbolagade, ES Adegoke | Brainstorming | 2014 | Discusses issues related to the use of the technique |
| Brainstorming a Backchat | D Somerville | Brainstorming | 2004 | Discusses issues related to the use of the technique |
| The Persistence of Brainstorming | MAM Gobble | Brainstorming | 2014 | Discusses issues related to the use of the technique |
| Brainstorming: Experiences from two thousand teams. | JD Antoszkiewicz | Brainstorming | 1992 | Discusses issues related to the use of the technique |
| Fifth Brainstorming Week on Membrane Computing | A Romero-Jiménez, A Riscos-Núñez | Brainstorming | 2007 | Discusses issues related to the use of the technique |
| IdeaGens: A Social Ideation System for Guided Crowd Brainstorming | J Chan, S Dang, P Kremer, L Guo | Brainstorming | 2014 | Discusses issues related to the use of the technique |
| DEVIL'S ADVOCACY IN MANAGERIAL DECISION-MAKING | CR Schwenk | Devils Advocacy | 1984 | Discusses issues related to the use of the technique |
| Socratic strategies and Devil's Advocacy in synchronous CMC debate | A Walker | Devils Advocacy | 2004 | Discusses issues related to the use of the technique |
| Diversity, eccentricity, and Devil's Advocacy | CR Schwenk | Devils Advocacy | 1998 | Discusses issues related to the use of the technique |
| The Red Teaming Essential | C Matherly | Red Teaming | 2013 | Discusses issues related to the use of the technique |
| Reflections from a Red Team Leader | Susan Craig | Red Teaming | 2007 | Discusses issues related to the use of the technique |
| The "Red Team" Forging a | Malone, Timothy G. and Reagan E. Schaupp | Red Teaming | 2002 | Discusses issues related to the use of the technique |
| Red Teaming for Law Enforcement | Michael Meeham | Red Teaming | 2007 | Discusses issues related to the use of the technique |

| | | | | |
|--|---|------------------------------|------|---|
| Perceptions of brainstorming in group | WC Rowatt, K Nesselroade | Brainstorming | 1997 | Discusses issues related to the use of the technique (perception of the technique) |
| Reflective Inquiry in the Virtual Brainstorming of Writing Tasks | CK Man | Brainstorming | nd | Discusses issues related to the use of the technique/Study lacks counterfactual condition |
| Drivers and outcomes of scenario planning | T J. Chermack, K Nimon | Alternative Futures Analysis | 2013 | Discusses issues related to the use of the technique--reasoning styles |
| Water Demand Projection in Distribution Systems Using a Novel Scenario Planning Approach | M Cabral, D Loureiro, A Mamade, D Covas | Alternative Futures Analysis | 2014 | Duplicates a previous observation |
| The comparative effectiveness of dialectical inquiry and Devil's Advocacy | DM Schweiger, PA Finger | Devils Advocacy | 1984 | Duplicates a previous observation |
| Re-examining the Team A-Team B exercise | RC Reich | Team A/Team B | 1989 | Duplicates a previous observation |
| TEAM B: THE TRILLION DOLLAR EXPERIMENT. | A Cahn | Team A/Team B | 1993 | Duplicates a previous observation |
| The Speaking Ability of the Eleventh Grade Students of SMA 1 Mejobo Kudus in the Academic Year 2011/2012 Taught by Using Round Robin Brainstorming | I Faizah | Brainstorming | 2012 | Educational outcome |
| THE ROLE OF BRAINSTORMING AND CLUSTERING IN WRITING A COMPOSITION | M SELVIANA | Brainstorming | 2007 | Educational outcome |
| The Use of Brainstorming to Improve the Writing Ability in Recount Text of the Tenth Grade | EW Suryani | Brainstorming | 2012 | Educational outcome |
| THE EFFECT OF APPLYING | M Fransisca, Z Zainuddin | Brainstorming | 2012 | Educational outcome |

| | | | | |
|---|---|---------------------------------|------|---|
| BRAINSTORMING TECHNIQUE ON THE | | | | |
| The Effect of Using Brainstorming Strategy in Developing Creative Problem Solving Skills | BA Al-khatib | Brainstorming | 2012 | Educational outcome |
| COMPARISON OF EFFECTIVENESS OF THE THREE METHODS OF BRAINSTORMING, SYNECTICS AND DEDUCTIVE ON INCREASING CREATIVE THOUGHT IN ... | | Brainstorming | 2010 | Educational outcome |
| ACH0: A Tool for Analyzing Competing Hypotheses | Lance Good, Jeff Shrager, Mark Stefik, Peter Pirolli, and Stuart Card | ACH | 2004 | Proposes an improvement or modification of the technique- but no evaluation |
| Deception Detection by Analysis of Competing Hypotheses | Frank J. Stech and Christopher Elsaesser | ACH | 2005 | Proposes an improvement or modification of the technique- but no evaluation |
| DECIDE™ Hypothesis Visualization Tool | Diane Cluxton and Stephen G. Eick | ACH | 2005 | Proposes an improvement or modification of the technique- but no evaluation |
| Extending Heuer's Analysis of Competing Hypotheses Method to Support Complex Decision Analysis | Marco Valtorta, Jiangbo Dang, Hrishikesh Goradia, Jingshan Huang, and Michael Huhns | ACH | 2005 | Proposes an improvement or modification of the technique- but no evaluation |
| Analysis of Competing Hypotheses using Subjective Logic | Simon Pope, Audun Jøsang | ACH | 2005 | Proposes an improvement or modification of the technique- but no evaluation |
| Transformative Scenario Planning | Adam Kahane | Alternative Futures Analysis | 2013 | Proposes an improvement or modification of the technique- but no evaluation |
| Scenario planning with integrated quantification | J Luis Cordeiro, S Hirsch, P Burggraf, C Daheim | Alternative Futures Analysis | 2013 | Proposes an improvement or modification of the technique- but no evaluation |

| | | | | |
|---|--|------------------------------|------|--|
| Normative scenario approach: a vehicle to connect adaptation planning and development needs in developing countries | L Bizikova, L Pintér, N Tubiello | Alternative Futures Analysis | 2014 | Proposes an improvement or modification of the technique-but no evaluation |
| Using Prospective Vision and Multi-Criteria Decision Analysis with Scenario Planning. | Carlos Gomes, Francisco Simoes, Helder Costa | Alternative Futures Analysis | nd | Proposes an improvement or modification of the technique-but no evaluation |
| Effective Scenario Composition for the Revelation of Blind Spots in Critical Infrastructure Protection Planning | WJ Tolone, SW Lee, WN Xiang, RK McNally | Alternative Futures Analysis | nd | Proposes an improvement or modification of the technique-but no evaluation |
| Brainstorming: Consensus Learning in Practice | D Plewczynski | Brainstorming | 2009 | Proposes an improvement or modification of the technique-but no evaluation |
| Take five for better brainstorming | S Thiagarajan | Brainstorming | 1991 | Proposes an improvement or modification of the technique-but no evaluation |
| Co-BrainSystem: Supporting brainstorming to enhance collaborative work in educational environments | J Herbert | Brainstorming | 2013 | Proposes an improvement or modification of the technique-but no evaluation |
| Synectics and brainstorming determinations of the ideative fluency of Nigerian adolescents | JO Akinboye | Brainstorming | 1980 | Relevant-but full text only in French (abstract was in English) |
| From Cultural Diversity To Group Creativity | HC Wang | Brainstorming | 2011 | Study is comparison of a variation of the technique |
| Idea generation in brainstorming and turn-taking groups | R Manning | Brainstorming | 1998 | Study is comparison of a variation of the technique |
| Acquisition of brainstorming behavior: differential effects of model status and symbolic | OH Revels | Brainstorming | 1980 | Study is comparison of a variation of the technique |
| Effects of Communication | A Miura | Brainstorming | 2003 | Study is comparison of a variation of the technique |

| | | | | |
|--|------------------------------------|---------------|------|---|
| Medium and Goal Setting on Group Brainstorming | | | | |
| Descriptive Evidence from Audit Practice on SAS No | JL Gissel, KM Johnstone | Brainstorming | 2007 | Study is comparison of a variation of the technique |
| Make Your Lecture Interesting with Discussion and Brainstorming Methods | JR Sonwane | Brainstorming | nd | Study is comparison of a variation of the technique |
| A Study of the Influence of Participants' Experience to the Brainstorming | CS HSU, CD CHEN, LJ HWANG, SA KAO | Brainstorming | nd | Study is comparison of a variation of the technique |
| The Future of Electronic Brainstorming | S Gauvin, MC Roy | Brainstorming | 1998 | Study is comparison of a variation of the technique |
| Personality, process, and performance in interactive brainstorming groups | AU Bolin, GA Neuman | Brainstorming | 2006 | Study is comparison of a variation of the technique |
| Idea production in nominal and virtual groups: does computer-mediated communication improve group brainstorming? | R Ziegler, M Diehl, G Zijlstra | Brainstorming | 2000 | Study is comparison of a variation of the technique |
| Process structuring in electronic brainstorming | AR Dennis, JS Valacich, T Connolly | Brainstorming | 1996 | Study is comparison of a variation of the technique |
| The remote associates test as a predictor of productivity in brainstorming groups | GB Forbach, RG Evans | Brainstorming | 1981 | Study is comparison of a variation of the technique |
| Making Group Brainstorming More Effective: Recommendations From an Associative Memory Perspective | Z Yun | Brainstorming | 2010 | Study is comparison of a variation of the technique |
| Interpersonal perception and group | FM Jablin, RL Sorenson, DR Seibold | Brainstorming | 1978 | Study is comparison of a variation of the technique |

| | | | | |
|--|---|---------------|------|---|
| brainstorming performance | | | | |
| Research frn Electronic Brainstorming | W Kanliang, X Youmin, W Yingluo | Brainstorming | 1994 | Study is comparison of a variation of the technique |
| Effects of gesture-based avatar-mediated communication on brainstorming and negotiation tasks | CS Ang, A Bobrowicz, P Siriaraya, J Trickey | Brainstorming | 2013 | Study is comparison of a variation of the technique |
| Effects of "brainstorming" instructions on creative problem solving by trained and untrained subject | SJ Parnes, A Meadow | Brainstorming | 1959 | Study is comparison of a variation of the technique |
| Creative supergroups: Group performance as a function of individual performance on brainstorming tasks | WK Graham, PC Dillion | Brainstorming | 1973 | Study is comparison of a variation of the technique |
| AUDIT PARTNER LEADERSHIP TONE AND PROFESSIONAL SKEPTICISM IN FRAUD BRAINSTORMING | S Dennis, KM Johnstone | Brainstorming | 2014 | Study is comparison of a variation of the technique |
| Groupgarden: supporting brainstorming through a metaphorical group mirror on table or wall | S Tausch, D Hausen, I Kosan, A Raltchev | Brainstorming | 2014 | Study is comparison of a variation of the technique |
| Asynchronous distance education forum-Brainstorming vs. Snowballing: a case study for | KPM Xenos | Brainstorming | 2009 | Study is comparison of a variation of the technique |
| An Audit Partner-led Field Intervention in Fraud Brainstorming | S Dennis, KM Johnstone | Brainstorming | 2014 | Study is comparison of a variation of the technique |

| | | | | |
|---|---|------------------------------|------|--|
| Participant-driven GSS: Quality of Brainstorming and Allocation of Participant Resources. | JH Helquist, EL Santanen, J Kruse | Brainstorming | 2007 | Study is comparison of a variation of the technique |
| Benchmarking bio-inspired designs with brainstorming in terms of novelty of design outcomes | S Keshwani, TA Lenau, SA Kristense | Brainstorming | 2013 | Study is comparison of a variation of the technique |
| Experiential effects of dialectical inquiry, Devil's Advocacy and consensus approaches | DM Schweiger, WR Sandberg | Devils Advocacy | 1989 | Study is comparison of a variation of the technique |
| Effects of Devil's Advocacy and dialectical inquiry on decision making | CR Schwenk | Devils Advocacy | 1990 | Study is comparison of a variation of the technique |
| Journ+A184:G184al Article | RL Priem, KH Price | Devils Advocacy | 1991 | Study is comparison of a variation of the technique |
| Structuring Conflict in Individual, Face-to-Face, and Computer-Mediated Group Decision Making | JS Valacich, C Schwenk | Devils Advocacy | 1985 | Study is comparison of a variation of the technique |
| Confirmation Bias in the Analysis of Remote Sensing Data | Paul Lehner, Leonard Adelman, Robert Distasio, Marie Erie, Janet Mittel, Sherry Olson | ACH | 2009 | Technique addressed in study was not one of those included in this study as defined in the Tradecraft Primer |
| SPURS: A framework towards Scenario Planning for Unexpected events, Response, and Startle using research, horror films, and video games | J BARNETT, BLW WONG, R ADDERLEY, M SMITH | Alternative Futures Analysis | 2013 | Technique addressed not one of the 6 in this study |
| Locating monitors in water distribution systems: Red team-blue team exercise | WM Grayman, A Ostfeld, E Salomons | Red Teaming | 2006 | Technique addressed not one of the 6 in this study |

| | | | | |
|--|---|-----------------|------|--|
| Straw persons and Devil's Advocacy | T Mehigan | Devils Advocacy | 1988 | Technique addressed not one of the 6 in this study |
| Deconstruction as Devil's Advocacy: A Shavian Alternative | RF Dietrich | Devils Advocacy | 1986 | Technique addressed not one of the 6 in this study |
| Royal tombs and preterhuman ancestors: a Devil's Advocacy | D Henige | Devils Advocacy | 1977 | Technique addressed not one of the 6 in this study |
| The Devil's Advocacy: When and why inhaled therapies for tuberculosis may not work | AB Yadav, AK Singh, RK Verma, M Mohan, AK Agrawal | Devils Advocacy | 2011 | Technique addressed not one of the 6 in this study |
| "Could God be Temporal?" A Devil's Advocacy | J King-Farlow | Devils Advocacy | 1963 | Technique addressed not one of the 6 in this study |
| B. Short description of the research projects of Team B | J SCHOUKENS, RIK PINTELON, Y ROLAIN | Team A/Team B | nd | Technique addressed not one of the 6 in this study |
| Final Report-Team B Navigation | J Moore, D Baird, K Sugimoto | Team A/Team B | 2003 | Technique addressed not one of the 6 in this study |
| Encouragement Through Community: An Ethnographic Study of Team B | L Reeves | Team A/Team B | 2015 | Technique addressed not one of the 6 in this study |
| Team B-BAM! | G Botteon, K Gursahaney | Team A/Team B | nd | Technique addressed not one of the 6 in this study |
| CS 692 RPE 7 Position Paper: Team B | K Cerar, AM Seto, J Zarnett | Team A/Team B | nd | Technique addressed not one of the 6 in this study |
| LibGuides Sandbox for Library Schools. 2014 SLIS 758 TEAM B. Home. | F TU, K Holder, E Evans | Team A/Team B | 2014 | Technique addressed not one of the 6 in this study |
| Jenna Cooley Forestry-Team B Samantha Holly | L Terango | Team A/Team B | nd | Technique addressed not one of the 6 in this study |
| ME 402-TTI Compressor-Team B | S Crouch, D Sharpe, C Odell, F Ragin | Team A/Team B | 2010 | Technique addressed not one of the 6 in this study |
| Cornwall Outdoors hosts team b | C Council | Team A/Team B | 2012 | Technique addressed not one of the 6 in this study |
| B. Short Description of the Research Projects of Team A | N DEBLAUWE, | Team A/Team B | | Technique addressed not one of the 6 in this study |

| | | | | |
|--|---|------------------------------|------|--|
| | LEO VAN BIESEN | | | |
| Team B | E Yamova - | Team A/Team B | 2014 | Technique addressed not one of the 6 in this study |
| Team B. Site within the Boundaries of Shchedrina, Sarafanovskaya, Detskaya, Napolnaya and Barrikad Streets | I Afanasieva | Team A/Team B | 2014 | Technique addressed not one of the 6 in this study |
| Energizing commitment to change in a team-building intervention: A FIRO-B approach | GH Varney, RJ Hunady | Team A/Team B | 1978 | Technique addressed not one of the 6 in this study |
| B-Human team description for robocup 2008 | T Röfer, T Laue, A Burchardt, E Damrose | Team A/Team B | 2008 | Technique addressed not one of the 6 in this study |
| Illusion of Devil's Advocacy: How the Justices of the Supreme Court Foreshadow Their Decisions during Oral Argument, The | S Schullman | Devils Advocacy | 2004 | Technique addressed not one of the 6 in this study (describes physical red teaming) |
| Second public hearing of the National Commission on Terrorist Attacks Upon the United States | Bogdan Dzakovic | Red Teaming | 2003 | Technique addressed not one of the 6 in this study (describes physical red teaming) |
| Planning of strategic innovation aimed at environmental sustainability | PJ Partidario, J Vergragt | Alternative Futures Analysis | 2002 | Technique addressed not one of the 6 in this study (describes statistical brainstorming, could not determine what this technique is) |
| Using Scenario Planning to Inform Pedagogical Practice in Virtual Worlds in Schools: Collaboration and Structure | CA Bonfield, KJ Burden | Alternative Futures Analysis | 2012 | Technique addressed not one of the 6 in this study (describes statistical brainstorming, could not determine what this technique is) |
| TRIGGERING STUDENT CREATIVITY THROUGH | J Solana-Gutierrez, MD Bejarano-Carrión | Brainstorming | 2012 | Technique addressed not one of the 6 in this study (describes statistical brainstorming, could |

| | | | | |
|--|--|---------------------------------|------|---|
| STATISTICAL BRAINSTORMING -AN | | | | not determine what this technique is) |
| The effects of scenario planning on perceptions of conversation quality and engagement | J Veliquette, LM Coons, SL Mace, T Coates | Alternative Futures Analysis | 2012 | Technique addressed not one of the 6 in this study--appears to be related to scenarios as a statistical tool |

Assessing Studies

The 46 evaluative studies were examined for reported effect of the technique and quality of evidence

Studies were assessed by examining the reported effect of the technique on rigor and accuracy. Each study was coded as followed and classified as effective or ineffective:

- | | | |
|-------------------|---|----------------------------|
| 1-Negative Effect | } | Considered "not effective" |
| 2-No Effect | | |
| 3-Mixed Effect | | |
| 4-Positive Effect | } | Considered "effective" |

Effectiveness was based on the reported impact of the technique on rigor and accuracy:

- Rigor : Analysis is rigorous when it is "in-depth" This criterion is called "coherence" because rigorous analysis coheres to a set of standards

Examples: questioning the reliability of sources, considering multiple hypotheses

- Accuracy : Analysis is accurate when there is a high degree of correspondence between the analytic judgment and what happened in the external world

Next each study was evaluated on the credibility of the evidence. Using a modified version of the Maryland Scientific Methods Scale. The modification was the addition of a category for meta-analyses:

More Credible
↓
Less Credible

| Level | Description | Example |
|-------|---|---|
| 1 | Correlation between a structured analytic technique and a measure of either rigor or accuracy measures | Analysts report using brainstorming generates more ideas |
| 2 | Measures of efficacy before and after using a technique, with no comparable control condition | Analysts generate more ideas after using brainstorming |
| 3 | Measures of efficacy before and after using a technique in experimental and comparable control conditions | Analysts generated more ideas in a treatment group but there was no improvement in efficacy in a control group |
| 4 | Measures of efficacy before and after the program in multiple experimental and control units, controlling for other variables that influence analytic performance | Analysts generated more ideas in a treatment group but there was no improvement in efficacy in a control group controlling for level of expertise and cognitive reasoning style |
| 5 | Random assignment of technique and control conditions to units | Analysts were randomized into a treatment group using techniques and a control group |
| 6 | Study is a systematic review | Multiple research reports of brainstorming were analyzed using meta-analysis |

10.0 METHODOLOGICAL APPENDIX C: EXPERIMENT PROCEDURES

Participant Selection

The study population was security and intelligence studies students at the Graduate School of Public and International Affairs (GSPIA) at the University of Pittsburgh. Participants were notified of the experiment via list-serv, solicitation during class time, and fliers posted around GSPIA. Upon completion of the experiment, participants received a small monetary incentive (\$20). Prior to the experiment, a pilot test was conducted with three participants to test and refine the experimental materials and procedures.

Experimental Task

21 participants volunteered for the experiment and were randomized into two groups: using Indicators or Analysis of Competing Hypotheses (ACH). These groups were further broken down into analytic “teams” of 3-4 students. Given the greater ease of using Indicators only the ACH group received instruction in how to use the technique. A 3 hour ACH workshop was held to cover the rationale and to practice using the technique:

Outline for Alternative Competing Hypotheses Training

- 1) Housekeeping: Questionnaire/Verbal consent form
- 2) Module 1: Logic of ACH
 - 1.1 What is ACH and why use it?
 - 1.2 Components of ACH
 - 1.3 Knowledge check #1
- 3) Module 2: ACH procedures
 - 2.1 How do you use ACH?
 - 2.2 PARC Software with a simple interactive example
 - 2.3 Knowledge check
- 4) Lunch break

5) Module 3: “Who killed Benazir Bhutto?”

Participants were given a questionnaire throughout the workshop to ensure they understood the rationale and how to use the technique. The class score was an 82 percent, suggesting a moderate level of comprehension.

Analytic teams were then given a date for their team to participate in the analytic experiment. During the three hour experiment, a facilitator (the author) used a protocol for Indicators and ACH. The following is the Indicators protocol:

Facilitator Guide: Indicators

1) **Introduction of Research:**

“Thank you for taking part in this research. Over the next 3 hours we will be conducting an analytic task simulating how analysis is conducted in the intelligence community. We will take a 10 minute break at the halfway point.

All data will be kept confidential on a password protected computer. Your participation is voluntary. Throughout the process I, the facilitator, will assist in structuring the conversation and managing time.

The purpose of this activity is to determine the likelihood Syria will meet the deadline set by Security Council Resolution 2118, for completed chemical weapon disarmament by June 30th, 2014. Currently, 30% of the weapons have been destroyed”

2) **Documents:** “Please look through these background articles over the next few minutes. [Distribute documents, hand out judgment sheet #1 after articles have been read then collect after each is filled out]

[Begin Recording]

4) **Generate:** “Now we will begin our analysis. We will start by determining the factors that could affect compliance with the chemical weapons agreement. [go around the room with each participant providing an idea, list all barriers on whiteboard, spend no more than 50 minutes]

5) **Refine:** “Now we will review all of our ideas and see if any can be combined or eliminated” [go through each factor and ask if it should be combined with another idea or eliminated, spend no more than 15min]

6) **Break:** [10 minute break]

7) **Ranking Importance:** “What barriers do you think are most important? Now we will group them into three categories: very important, somewhat important, and not important. [Spend no more than 30 minutes ranking each item by going around the room and asking for consensus]

8) **Assessing Likelihood:** “Now we will attempt to determine which of the most likely factors: very likely, somewhat unlikely, and not likely. [Spend no more than 30 minutes ranking each item by going around the room and asking for consensus]

9) **Final assessment:** “Please take a few minutes to record what you think will be the outcome of the Syrian chemical weapons agreement.” [Pass out judgment sheets]

The ACH protocol differs slightly:

Facilitator guide- Analysis of Competing Hypotheses

1) **Introduction of Research:**

“Thank you for taking part in this research. Over the next 3 hours we will be conducting an analytic task simulating how analysis is conducted in the intelligence community. We will take a 10 minute break at the halfway point.

All data collected will be kept confidential on a password protected computer. Your participation is voluntary. Throughout the process I, the facilitator, will assist in structuring the conversation and managing time.

The purpose of this activity is to determine the likelihood Syria will meet the deadline set by Security Council Resolution 2118, for completed chemical weapon disarmament by June 30th, 2014.

2) **Documents:** “Please look through these background articles over the next 10 minutes. [distribute questionnaire, documents, open PARC software, hand out judgment sheet 1 after articles have been read then collect after each is filled out]

[Begin recording]

3) **Begin Analysis:** “Now we will begin with the analysis. We will begin by listing the simple hypotheses” [ask for simple hypotheses, should take no more than a minute]

4) **Evidence:** What pieces of evidence would confirm or disconfirm these two hypotheses? Remember you may use evidence from your own experience and/or the provided documents.” [Go around room gathering evidence, write on board, spend no more than 20 minutes listing evidence]

5) **Refine Evidence:** “Now we will review all of our pieces of evidence and see if any can be combined or eliminated” [go through each factor and ask if it should be combined with another piece of evidence or eliminated, spend no more than 15min]

6) **Break:** [10 minute break]

7) **Assess Diagnosticity:** “Now we will assess the diagnosticity of each hypothesis. Where there is disagreement we will take a simple vote to determine the diagnosticity and a flag will be placed to indicate disagreement. [Begin, at upper left cell and work through each hypothesis. Spend no more than 50 minutes assessing diagnosticity]

8) **Assess Sensitivity:** “Consider each piece of evidence. Are there any pieces of evidence that if altered would dramatically change the result?

9) **Assess result:** “What are/is the most credible hypotheses/hypothesis? Could we eliminate any/one?”

10) **Final assessment:** “Please take a few minutes to record what you think will be the outcome of the Syrian chemical weapons agreement.” [Pass out judgment sheets, spend no more than 10 minutes]

At the beginning of the simulation, participants were presented with a map of Syria dated two months prior which included the rebel positions and chemical weapons sites, a timeline of key events, and the UN resolution. After examining these documents for 10 minutes, participants filled out a judgment sheet and then the same judgment sheet after using ACH or Indicators:

Name: _____ Date: _____

1) What percentage of Syrian government-held chemical weapons will be destroyed by June 30th? (Currently, 30% of the weapons have been destroyed)

_____ %

2) What is your confidence your judgment for #1? (circle)

High – Moderate – Low

3) What do you think the likelihood is that virtually all of Syrian government-held chemical weapons will be destroyed by the deadline, June 30th, 2014? (circle)

Almost Certainly – Probably/Likely – Even Chance – Unlikely – Remote

4) Provide a rationale for your answer for #3:

The rationales were next coded. Below is an example of how data was coded. Each narrative was analyzed for any hypotheses that the participant cited that might have an impact on the outcome of the disarmament process:

| Respondent ID | Narrative | Hypotheses |
|---------------|---|---|
| 2001 | Based on the evidence I saw in the previous papers handed out, I believe the Syrian government is in no rush to destroy all the chemical weapons, [indiscernable] incompetent and slow. As long as they are [indiscernable] will not be [indiscernable] | -Intentional slowdown-Syria |
| 2002 | The deadline has been pushed back repeatedly [sic], and leaders on the outside--Russia and western nations-- are collaborating to end the problem; limited transparency. I think Syria's chemical weapons program will be eradicated, but will happen after 6/30/2014 | -International cooperation- the West -Limited transparency |
| 2003 | To this point, the process has not remained on schedule-if the destruction process takes 90 days after removal, they would need to remove an additional 60% immediately to be successful | -Insufficient time to reach goal |

11.0 BIBLIOGRAPHY

"Aggregative Contingent Estimation (ACE)" IARPA, accessed 26 April, 2015. Available at: <http://www.iarpa.gov/index.php/research-programs/ace>

Aid, Matthew. "The NSA's Data Haul Is Bigger Than You Can Possibly Imagine." *Foreign Policy*. August 13, 2013, available at: http://www.foreignpolicy.com/articles/2013/08/15/the_nsas_data_haul_is_bigger_than_you_can_possibly_imagine

Allison, Graham T. *Essence of Decision*. Boston: Little, Brown, 1971.

Arkes, Hal R. and James Kajdasz, "Intuitive Theories of Behavior" in *Intelligence Analysis: Behavioral and Social Scientific Foundations*, Baruch Fischhoff and Cherie Chauvin (eds.), Washington, D.C.: National Academies Press, 2011

Armstrong, Jon Scott, ed. *Principles of forecasting: a handbook for researchers and practitioners*. Vol. 30. Springer Science and Business Media, 2001.

Armstrong, J. Scott. "Findings from evidence-based forecasting: Methods for reducing forecast error." *International Journal of Forecasting* Volume 22, Issue 3, 2006, pp. 583-598.

Armstrong, J. Scott. "How to make better forecasts and decisions: Avoid face-to-face meetings." *Foresight: The International Journal of Applied Forecasting* 5 (2006): 3-15.

Bar-Joseph, Uri, and Rose McDermott. "Change the analyst and not the system: A different approach to intelligence reform." *Foreign Policy Analysis* Volume 4, Issue 2, 2008, pp. 127-145.

Barnett, Thomas. "Just How Intelligent is the National Intelligence Council's Global Trends 2030?" *Time*, December 21, 2012. Available at: <http://nation.time.com/2012/12/21/just-how-intelligent-is-the-national-intelligence-councils-global-trends-2030/>

Barki, Henri, and Alain Pinsonneault. "Small Group Brainstorming and Idea Quality Is Electronic Brainstorming the Most Effective Approach?." *Small Group Research*, Volume 32, Issue 2, 2001, pp. 158-205.

Beel, Jöran, and Bela Gipp. "Google Scholar's ranking algorithm: an introductory overview." *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*. Volume 1, 2009.

Bernard, H.R.. *Research Methods in Anthropology: Qualitative and Quantitative methods* (4th ed). Walnut Creek, California: AltaMira Press, 2006.

Betts, Richard K. *Enemies of Intelligence*. New York: Columbia University Press, 2007.

Billman, Dorrit, Gregorio Convertino, Jeff Shrager, P. Pirolli, and J. Massar. "Collaborative intelligence analysis with CACHE and its effects on information gathering and cognitive bias." In *Human Computer Interaction Consortium Workshop*. 2006.

Blau, Peter Michael. *Exchange and Power in Social Life*. Transaction Publishers, 1964.

Blomstrom, Sally, F. J. Boster, K. J. Levine, E. M. J. Butler, and S. L. Levine. "The effects of training on brainstorming." *Journal of the Communication, Speech and Theatre Association of North Dakota* (2000): 41.

Brasfield, Andrew D. *Forecasting Accuracy and Cognitive Bias in the Analysis of Competing Hypotheses*. Diss. Mercyhurst College, 2009.

Brunswik, Egon. *The Conceptual Framework of Psychology*. Vol. 1. No. 10. Univ of Chicago Pr, 1952.

Campbell Collaboration. "About Us" Accessed 22 October, 2014. Available at: http://www.campbellcollaboration.org/about_us/index.php

Campbell, Donald T. "The informant in quantitative research." *American Journal of Sociology*, 1955, pp. 339-342.

Campbell, Donald. Descriptive Epistemology: Psychological, Sociological, and Evolutionary." William James Lectures, Harvard University. Reprinted as pp. 435-86 in *Methodology and Epistemology for Social Science: Selected Papers*, edited by E. Samuel Overman. Chicago, IL: The University of Chicago Press (1977).

Campbell, Donald T., and Donald W. Fiske. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin*, Volume 56, Issue 2, 1959.

Campbell, Donald Thomas, Julian C. Stanley, and Nathaniel Lees Gage. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin, 1963.

Campbell, Donald T., Richard D. Schwartz, and Lee Sechrest. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Vol. 111. Chicago: Rand McNally, 1966.

Chamberlin, Thomas Chowder. "The method of multiple working hypotheses." *Science* Volume 15, Issue 366, 1890, pp. 92-96.

Chase, William G., and Herbert A. Simon. "Perception in chess." *Cognitive Psychology*, Volume 4, Issue 1, 1973, 55-81.

Chauvin, Cherie, and Baruch Fischhoff, eds. *Intelligence Analysis: Behavioral and Social Scientific Foundations*. National Academies Press, 2011.

Cheikes, Brant A., Mark J. Brown, Paul E. Lehner, and Leonard Adelman. "Confirmation bias in complex analyses." *MITRE Center for Integrated Intelligence Systems*, 2004.

Chen, Xiaotian. "Google Scholar's Dramatic Coverage Improvement Five Years after Debut a." *Serials Review* 36.4 (2010): 221-226.

Churchman, C. West. *Design of Inquiring Systems: Basic Concepts of Systems and Organization*. New York: Basic Books (1971).

Cochrane Collaboration, Cochrane Glossary, Accessed 22 October, 2014. Available at: <http://www.cochrane.org/glossary>

Cochrane Collaboration. "About Us." Accessed 10 October, 2014. Available at: <http://www.cochrane.org/about-us>

Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction. *Final Report of the Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction*. Washington: GPO, 2005.

Cook, Thomas D., and Donald Thomas Campbell. *Quasi-experimentation: Design and Analysis for Field Settings*. Rand McNally, 1979.

Cook, T.D. "Postpositivist critical multiplism." In *Social science and Social Policy* R.L. Shotland and M. M. Mark (Eds.), (pp. 21-62). Beverly Hills, CA: Sage (1985).

Cooper, Harris. *Research Synthesis and Meta-analysis: A Step-by-Step approach*. Vol. 2. Sage Publications, 2009.

Coulthart, Stephen and Jay Rickabaugh. "[Bracketing] the Black Swan (Part II)." 25 June, 2014, available at: <http://research.ridgway.pitt.edu/blog/2014/06/25/bracketing-the-black-swan-part-ii/>

Creswell, John W. *Research design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage, 2013.

Crosston, Matthew and Stephen Coulthart. "Analyzing Post 9/11 Intelligence Degree Programs: Failure to Launch or Cause for Hope?" *Intelligence and National Security* (under review).

Davies, Huw T. O., Sandra M. Nutley, and Peter C. Smith. *What Works?: Evidence-Based Policy and Practice in Public Services*. Policy Press, 2000.

Davis, Jack. "Combatting Mindset." *Studies in Intelligence* 35 (1992): 13-18.

Davis, Jack. *A Compendium of Analytic Tradecraft Notes*. Central Intelligence Agency, Directorate of Intelligence, 1997.

Davis, Jack. "Improving Intelligence Analysis at CIA: Dick Heuer's Contribution to Intelligence Analysis" in *Psychology of Intelligence Analysis*. Heuer, Richards J. (author), Washington D.C.: Center for the Study of Intelligence Analysis, 1999.

Davis, Jack. *Improving CIA Analytic Performance: Strategic Warning*. Central Intelligence Agency, Washington DC, 2002.

Defense Intelligence Agency. "Graduating 4,000 Analysts from Analysis 101." December 2011, available at: <http://www.dia.mil/public-affairs/news/2011-12-19.html>

Defense Intelligence Agency, "IC Centers for Academic Excellence" accessed February 7, 2014, available at: <http://www.dia.mil/training/iccentersforacademicexcellence.aspx>

Denzin, N.K. . *The Research Act: A Theoretical Orientation to Sociological Methods* (2nd ed.). New York: McGraw-Hill (1978).

Diesing, Paul. *How Does Social Science Work?: Reflections on Practice*. University of Pittsburgh Press, 1991.

Dillman, Don A. *Mail and Internet surveys: The Tailored Design Method--2007 Update with new Internet, Visual, and Mixed-mode Guide*. John Wiley & Sons, 2011.

Dixon, Nancy M., and Laura A. McNamara. "Our experience with Intellipedia: An ethnographic study at the Defense Intelligence Agency." Retrieved August 30 2008, 2011.

Dugosh, Karen Leggett, Paul B. Paulus, Evelyn J. Roland, and Huei-Chuan Yang. "Cognitive stimulation in brainstorming." *Journal of Personality and Social Psychology* Volume 79, Issue 5, 2000.

Dunn, William N. "Pragmatic Eliminative Induction: Proximal Range and Context validation in applied social experimentation." *Philosophica*, volume 60, 1997, 75-112.

Dunn, William N. "Using the Method of Context Validation to Mitigate Type III errors in Environmental Policy Analysis." *Knowledge, Power and Participation in Environmental Policy Analysis*, 2001, pp. 417-436.

Dunn, William N. *Public Policy Analysis: An Introduction*. Upper Saddle River, NJ: Pearson, 2008/2012

Dunn, William N., and Fredric W. Swierczek. "Planned Organizational Change: Toward Grounded Theory." *The Journal of Applied Behavioral Science*, Volume 13, Issue 2, 1977, pp. 135-157.

Easterbrook, P.J, R Gopalan, J.A Berlin, D.R Matthews "Publication Bias in Clinical Research." *The Lancet* 337.8746 (1991): 867-872.

Feder, Stanley. "Factions and policon: New ways to analyze politics." *Inside CIA's Private World: Declassified Articles from the Agency's Internal Journal* 1992 (1955): 274-292.

Fingar, Thomas. *Remarks and QandA by the Deputy Director of National Intelligence For Analysis and Chairman*. National Intelligence Council presented at 2008 INSA Analytic Transformation Conference. Orlando, Florida. September 4, 2008

Fishbein, Warren, and Gregory Treverton. *Rethinking" Alternative Analysis" to Address Transnational Threats*. Washington D.C.: Central Intelligence Agency (2004).

Folker Jr, Robert D. *Intelligence Analysis in Theater Joint Intelligence Centers: An Experiment in Applying Structured Methods*. Washington D.C.: Joint Military Intelligence College, Center for Strategic Intelligence Research, 2000.

Garst, Ronald D., 2d ed. (Washington, DC: Defense Intelligence College, January 1989), 1; *Central Intelligence Agency, A Consumer's Guide to Intelligence* (Washington, DC: Public Affairs Staff, July 1995, 5-7.

Gehanno, Jean-François, Laetitia Rollin, and Stefan Darmoni. "Is the coverage of Google Scholar Enough to be used Alone for Systematic Reviews?" *BMC Medical Informatics and Decision Making*, Volume 13, Issue 1, 2013.

George, A. L. The case for multiple advocacy in making foreign policy. *American Political Science Review*, 66 (September 1975), 751-785.

van Gelder, Tim, "What is Argument Mapping?" 17 February 2009, available at: <http://timvangelder.com/2009/02/17/what-is-argument-mapping/>

Gladstone, Rick. "Syria Deaths Hit New High in 2014, Observer Group Says." *The New York Times*. January 1st, 2015. Available at: <http://www.nytimes.com/2015/01/02/world/middleeast/syrian-civil-war-2014-deadliest-so-far.html>

Gladwell, Malcolm. *Blink: The Power of Thinking Without Thinking*. Hachette Digital, Inc., 2007.

Glaser, Barney G. *Theoretical sensitivity: Advances in the methodology of grounded theory*. Vol. 2. Mill Valley, CA: Sociology Press, 1978.

Glaser, Barney G., and Anselm L. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Adlin, 1967.

"Global Agenda: Red Cell Intelligence analyst describes his role as 'devil's advocate' in the CIA," *University of Delaware Daily*, April 5, 2012, available at: <http://www.udel.edu/udaily/2012/apr/global-agenda-red-cell-040512.html>

Google Scholar, "About Google Scholar." Accessed 10 October, 2014. Available at: <http://scholar.google.com/intl/en/scholar/about.html>

Grabo, Cynthia M., and Jan Goldman. *Anticipating Surprise: Analysis for Strategic Warning*. University Press of America, 2004.

Graduate School of Public and International Affairs. "Major in Security and Intelligence Studies." Accessed 10 October, 2014. Available at: <https://www.gspia.pitt.edu/Academics/DegreePrograms/MasterofPublicInternationalAffairs/MajorInSecurityIntelligenceStudies/tabid/95/Default.aspx>

Greitzer, Frank L., and Kelcy Allwein. "Metrics and measures for intelligence analysis task difficulty." Panel Session, 2005 International Conference on Intelligence Analysis Methods and Tools. Vienna, VA. 2005.

Hammond, Kenneth R. *Human Judgment and Social Policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford University Press, 1996.

Helmer, Olaf, and Nicholas Rescher. "On the Epistemology of the Inexact Sciences." *Management science* 6.1 (1959): 25-52.

Heuer, Richards. *Quantitative Approaches to Intelligence Analysis*. Boulder, CO: Westview Press Inc, 1978.

Heuer, Richards J. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence Analysis, 1999.

Heuer, Richards J., and Randolph H. Pherson. *Structured Analytic Techniques for Intelligence Analysis*. CQ Press, 2010/ 2014

Heuer, Richards. "The Evolution of Structured Analytic Techniques." Presentation to the National Academy of Science, National Research Council Committee on Behavioral and Social Science Research to Improve Intelligence Analysis for National Security, Washington, DC, December 8, 2009.

Hollands, Justin G., and Christopher D. Wickens. "Engineering Psychology and Human Performance." *Journal of Surgical Oncology*, 1999.

Immerman, Richard H. "Transforming Analysis: The Intelligence Community's Best Kept Secret." *Intelligence and National Security*, Volume 26, 2011, pp. 159-181.

Janis, Irving L. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Oxford, England: Houghton Mifflin, 1972.

Jervis, Robert. *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press, 1976.

Jervis, Robert. "Reports, politics, and intelligence failures: The Case of Iraq." *The Journal of Strategic Studies*, Volume 29, Issue 1, 2006, 3-52.

Johnson, Rob. *Developing a Taxonomy of Intelligence Analysis Variables*. Washington, DC: The Center for the Study of Intelligence, 2002.

Johnston, Rob. *Analytic culture in the US intelligence community: An Ethnographic Study*. Central Intelligence Agency, Washington DC: Center for the Study of Intelligence, 2005.

Kahan, Dan M., et al. "Fear of democracy: A Cultural Evaluation of Sunstein on Risk." *Harvard Law Review* (2006): 1071-1109.

Kahan, Daniel, "System 1" and "System 2" are intuitively appealing but don't make sense on reflection: Dual process reasoning and science communication part 1" *The Cultural Cognition Project at Yale Law School*, 19 July 2013, available at: <http://www.culturalcognition.net/blog/2013/7/19/system-1-and-system-2-are-intuitively-appealing-but-dont-mak.html>

Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux 2011.

Kahneman, Daniel, and Gary Klein. "Conditions for Intuitive Expertise: A Failure to Disagree." *American Psychologist* 64.6 (2009): 515.

Kaplan, Abraham. *The conduct of inquiry*. Transaction Publishers, 1973.

Karelaia, Natalia, and Robin M. Hogarth. "Determinants of Linear Judgment: A Meta-Analysis of Lens Model Studies." *Psychological Bulletin*, Volume 134, Issue 3, 2008.

Kelly, Mary. "Intelligence Community Unites for 'Analysis 101'" *Morning Edition, National Public Radio*, May 07, 2007. Available at: <http://www.npr.org/templates/story/story.php?storyId=10040625>

Kenney, Michael. "Learning from the "Dark Side" – Identifying, Accessing and Interviewing Illicit Non-state Actors" in *Conducting Terrorism Field Research: A Guide*. Dolnik, Adam, ed., New York: Routledge, 2013.

Kent, Sherman. *Strategic Intelligence for American World Policy*. Princeton, NJ: Princeton University Press, 1949.

Kerbel, Josh. "Lost for words: The Intelligence Community's Struggle to Find its Voice." *Parameters*, Volume 38, Issue 2, 2008, pp. 102-112.

Kerlinger, F. N., and H. B. Lee. *Foundations of Behavioral Research*. Stamford: Wadsworth, 2000.

Khalsa, Sundri. "The Intelligence Community Debate over Intuition versus Structured Technique: Implications for Improving Intelligence Warning." *Journal of Conflict Studies* Volume 29, 2009.

Kohn, Nicholas William. *An Examination of Fixation in Brainstorming*. Diss. Texas A&M University, 2008.

Krizan, Lisa. *Intelligence essentials for everyone*. Washington D.C.: Joint Military Intelligence College, Center for Strategic Intelligence Research, 2000.

Kruglanski, Arie W. *The Psychology of Closed mindedness*. New York: Psychology Press, 2004.

Laipson, Eileen. "The Robb-Silberman Report, Intelligence, and Nonproliferation." *Arms Control Today*, Volume 35 Issue 5, 2005.

Lee, Thomas W. *Using Qualitative Methods in Organizational Research*. New York: Sage, 1999.

Light, Richard, and David B. Pillemer. *Summing Up*. Boston, MA: Harvard University Press, 1984.

Lincoln, Y. S., & Guba, E. G. *Naturalistic inquiry*. Newbury Park, CA: Sage, 1985.

Lindgren, Henry Clay, and Fredrica Lindgren. "Creativity, Brainstorming, and Orneriness: a Cross-Cultural Study." *The Journal of Social Psychology*, Volume 67, Issue.1, 1965, pp. 23-30.

Linstone, Harold A., and Murray Turoff, eds. *The Delphi Method: Techniques and Applications*. Vol. 29. Reading, MA: Addison-Wesley, 1975.

Linstone, Harold A., and Arnold J. Meltsner. *Multiple Perspectives for Decision Making: Bridging the Gap Between Analysis and Action*. New York: North-Holland, 1984.

Lockheed Martin, "Structured Analytic Techniques," available at: <http://www.lockheedmartin.com/us/products/intelligence-analysis-training/intelligence-training/counterterrorism-analysis.html>

Lucas, William A. *The Case Survey Method: Aggregating Case Experience*. No. 1515. Santa Monica, CA: Rand, 1974.

Lum, Cynthia, Leslie W. Kennedy, and Alison Sherley. "Are Counter-Terrorism Strategies Effective? The Results of the Campbell Systematic Review on Counter-terrorism Evaluation Research." *Journal of Experimental Criminology* Volume 2, Issue 4, 2006, pp.489-516.

MacEachin, Douglas J. *The Tradecraft of Analysis: Challenge and Change in the CIA*. Consortium for the Study of Intelligence, 1994.

Mandel, David R. "Canadian perspectives: Applied behavioral science in support of intelligence analysis." *Invited paper presented at the Public Workshop of the National*

Research Council Committee on Behavioral and Social Science Research to Improve Intelligence Analysis for National Security. Washington, DC: The National Academies. 2009.

Mangio, C.A. and Wilkinson, B.J. *State of Knowledge Relative to Intelligence Analysis. Volume 1: Cognitive Challenges for Intelligence Analysis - Past, Present, and Future*, AFRL-RH-WP-TR-2011-0049, 2011.

von Maravic, Patrick. "Limits of Knowing or the Consequences of Difficult-Access Problems for Multi-method Research and Public Policy." *Policy Sciences*, Volume 45, Issue 2, 2012, pp.153-168.

Marrin, Stephen. "CIA's Kent School: Improving Training for New Analysts." *International Journal of Intelligence and CounterIntelligence*, Volume 16, Issue 4, 2003, pp. 609-637.

Marrin, Stephen. "Intelligence Analysis: Structured Methods or Intuition?." *American Intelligence Journal*, Volume 25, Issue 1, 2007, pp. 7-16.

Marrin, Stephen. "Evaluating the Quality of Intelligence Analysis: By What (Mis) Measure?." *Intelligence and National Security*, Volume 27, Issue 6, 2012, pp. 896-912.

Martin, Kirsty, Mark Kebbell, Louise Porter, and Michael Townsley, "The Paradox of Intuitive Analysis and the Implications for Professionalism." *Journal of the AIPIO*, Volume 19, Issue 1, 2011.

McClelland, G.H. "Use of Signal Detection Theory as a Tool for Enhancing Performance and Evaluating Tradecraft in Intelligence Analysis," in *Intelligence Analysis: Behavioral and Social Scientific Foundations*, eds. B. Fischhoff and C. Chauvin, Washington, D.C.: National Academies Press, 2011.

Medina, Carmen. "The New Analysis." In *Analyzing Intelligence: Origins, Obstacles, and Innovations*, eds. R. George and J. Bruce, Washington DC: Georgetown University Press, 2008.

Miller, George A. "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information." *Psychological Review*, Volume 63, Issue 2, 1956.

Mitchell, Gordon R. "Team B intelligence coups." *Quarterly Journal of Speech*, Volume 92, Issue 2, 2006, pp. 144-173.

Mitroff, Ian I., and Murray Turoff. "Technological forecasting and assessment: science and/or mythology?." *Technological Forecasting and Social Change*, Volume 5, Issue 2 (1973), pp. 113-134.

Mitroff, Ian, I., V.P. Barabba, R.H. Kilmann. "The Application of Behavioral and Philosophical Techniques to Strategic Planning: A Case Study of a Large Federal Agency." *Management Science*, Volume 23. Pp. 44-58

Mitroff, Ian I., and James R. Emshoff. "On strategic assumption-making: A dialectical approach to policy and planning." *Academy of Management Review*, Volume 4, Issue 1 (1979), pp. 1-12.

Moore, David and Robert Hoffman "How might critical thinking and structured analytic techniques improve intelligence? Or..." unpublished manuscript (undated)

Moore, David T. *Critical Thinking and Intelligence Analysis*, Washington, D.C: National Defense Intelligence College, 2007.

Muller Jr, David G. "Intelligence analysis in red and blue." *International Journal of Intelligence and Counterintelligence*, Volume 21, Issue 1, 2007, pp.1-12.

Murrell, Audrey J., Alice C. Stewart, and Brent T. Engel. "Consensus Versus Devil's Advocacy: The Influence of Decision Process and Task Structure on Strategic Decision Making." *Journal of Business Communication*, Volume 30, Issue 4, 1993, pp. 399-414.

Myers, Isabel Briggs, *Intro to Type: A Guide to Understanding Your Results on the Myers-Briggs Type Indicator*, Palo Alto, CA: Consulting Psychologists Press, Inc., 1998..

Nikitin, Mary Beth D., Paul K. Kerr, Andrew Feickert. "Syria's Chemical Weapons: Issues for Congress." *Congressional Research Service*, September 30, 2013, available at: <http://www.fas.org/sgp/crs/nuke/R42848.pdf>

Nisbett, Richard E. *Human interference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ Prentice-Hall, 1980.

Office of the Director of National Intelligence, *ODNI Progress Report - WMD Commission Recommendations*. Washington, D.C., 2006.

Office of the Director of National Intelligence. "Intelligence Community Directive 206" Washington, D.C., 2007

Office of the Director of National Intelligence. "Intelligence Community Directive 207." Washington, D.C., 2007.

Offner, Anne K., Thomas J. Kramer, and Joel P. Winter. "The Effects of Facilitation, Recording, and Pauses on Group Brainstorming." *Small Group Research* Volume 27, Issue. 2, 1996, pp. 283-298.

O'Leary, Michael K., et al. "The Quest for Relevance: Quantitative International Relations Research and Government Foreign Affairs Analysis." *International Studies Quarterly*, 1974, pp. 211-237.

Organisation for the Prohibition of Chemical Weapons, "8% of Syrian Chemicals Still Remain to be Removed; Fact-Finding Mission in Syria; Some Progress on Syrian Production Facilities," 17 June 2014, available at: <http://www.opcw.org/news/article/8-of-syrian-chemicals-still-remain-to-be-removed-fact-finding-mission-in-syria-some-progress-on-s/>

Paulus, P. B., and Dzindolet, M. T., Social Influence Processes in Group Brainstorming. *Journal of Personality and Social Psychology*, Volume 64, 1993, pp. 575-586

Petroski, Henry. *Success through Failure: The Paradox of Design*. Princeton, NJ: Princeton University Press, 2006.

Pirolli, Peter, and Stuart Card. "The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis." *Proceedings of International Conference on Intelligence Analysis*. Vol. 5. McLean, VA: Mitre, 2005.

Platt, John R. "Strong Inference." *Science*, Volume 146, Issue 3642, 1964, pp. 347-353.

Pool, Robert, ed. *Field Evaluation in the Intelligence and Counterintelligence Context:: Workshop Summary*. National Academies Press, 2010.

Popper, Karl R. *The Logic of Scientific Discovery*. London: Hutchinson (1959).

Popper, Karl. *Objective Knowledge: An Evolutionary Approach*. Oxford: Oxford Clarendon Press, 1972.

Popper, Karl. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Routledge, 2014.

Priest, Dana, and William M. Arkin. *Top Secret America: The Rise of the New American Security State*. Hachette Digital, Inc., 2011.

Quaddus, M. A., Lai Lai Tung, L. Chin, P. P. Seow, and G. C. Tan. "Non-networked group decision support system: effects of Devil's Advocacy and dialectical inquiry." In

System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conference on, volume. 1, , 1998, pp. 38-47.

QSR International 2013, "NVivo 10 for Windows." Accessed 10 October, 2014. Available at: http://www.qsrinternational.com/products_nvivo.aspx

Rescher, Nicholas. *Predicting the Future: An Introduction to the Theory of Forecasting*. Albany, NY: SUNY press, 1998.

Rieber, Steven. "Intelligence Analysis and Judgmental Calibration." *International Journal of Intelligence and CounterIntelligence*, Volume 17, Issue 1, 2004, pp. 97-112.

Rieber, Steven and Neil Thomason, "Creation of a National Institute for Analytic Methods," *Studies in Intelligence*, 2007 at <https://www.cia.gov/library/center-for-the-study-of-intelligence/sci-publications/sci-studies>

Rood, Justin, "Analyze This: Inside the one spy agency that got pre-war intelligence on Iraq--and much else--right." *Washington Monthly*, January/February 2005.

Rosenthal, Robert, and Donald B. Rubin. "A Simple, General Purpose Display of Magnitude of Experimental Effect." *Journal of Educational Psychology*, Volume 74, Issue 2, 1982.

Rousseau, Denise M., ed. *The Oxford Handbook of Evidence-based Management*, Oxford: Oxford University Press, 2012.

Rudner, Martin. "Intelligence Studies in Higher Education: Capacity-Building to Meet Societal Demand," *International Journal of Intelligence and CounterIntelligence*, Volume 22, 2009.

Sackett, David L. *Evidence-based Medicine*. John Wiley & Sons, Ltd, 2000.

Scholtz, Jean, Emile Morse, and Tom Hewett. "In Depth Observational Studies of Professional Intelligence Analysts." *Human Performance, Situation Awareness and Automation Technology Conference*. 2004

Schwartz, Peter. *The Art of the Long View: Paths to Strategic Insight for Yourself and Your Company*. New York: Doubleday, 1991.

Schweiger, David M., William R. Sandberg, and James W. Ragan. "Group approaches for improving strategic decision making: A comparative analysis of dialectical inquiry, Devil's Advocacy, and consensus." *Academy of management Journal*, Volume 29, Issue 1, 1986, 51-71.

Schweiger, David M., William R. Sandberg, and Paula Rechner. "A Longitudinal Comparative Analysis of Dialectical Inquiry, Devil's Advocacy and Consensus Approaches to Strategic Decision Making." *Academy of Management Proceedings*. Vol. 1988. No. 1. Academy of Management, 1988.

Schwenk, Charles R. "Devil's Advocacy and Dialectical Inquiry Effects on Prediction Performance: Task Involvement As a Mediating Variable." *Decision Sciences*, Volume 15, Issue 4, 1984, pp. 449-462.

Schwenk, Charles R. "Effects of Devil's Advocacy on escalating commitment." *Human Relations*, Volume 41, Issue 10, 1988, pp. 769-782.

Schwenk, Charles. "A Meta-Analysis on the Comparative Effectiveness of Devil's Advocacy and Dialectical Inquiry." *Strategic Management Journal*, Volume 10, Issue 3, 1989, pp. 303-306.

Select Committee on Intelligence, U.S. Senate. *Recommendations of the Jeremiah Report*. Washington DC: GPO, 1998.

Shadish, William R., Thomas D. Cook, and Donald Thomas Campbell. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Houghton Mifflin, 2002.

Shanteau, James. "Competence in experts: The Role of Task Characteristics." *Organizational Behavior and Human Decision Processes*, Volume 53, Issue 2, 1992, pp. 252-266.

Shanteau, James. "Psychological Characteristics of Expert Decision Makers." *Expert Judgment and Expert Systems*. Springer Berlin Heidelberg, 1987. 289-304.

Sherman, Lawrence W. *Evidence-based Policing*. Washington, DC: Police Foundation, 1998.

Sherman, Lawrence W., et al. "Preventing Crime: What Works, What Doesn't, What's Promising. Research in Brief. National Institute of Justice." (1998).

Simon, Herbert. *Models of Man: Social and Rational; Mathematical Essays on Rational Human Behavior in Society Setting*. Wiley, 1957.

Simon, Herbert Alexander. *Administrative behavior*. Vol. 4. New York: Free Press, 1965.

Simon, Herbert A. "How big is a chunk." *Science*, Volume 183, Issue 4124., 1974, pp. 482-488.

Simon, Herbert A. "Theories of Bounded Rationality." *Decision and Organization*, Volume 1, 1972, pp. 161-176.

Spiegel, Alix. "So You Think You're Smarter Than A CIA Agent?" *National Public Radio*, April 02, 2014. Available at: <http://www.npr.org/blogs/parallels/2014/04/02/297839429/-so-you-think-youre-smarter-than-a-cia-agent>

Spracher, William C. National security intelligence professional education: A map of US civilian university programs and competencies. PhD Dissertation, The George Washington University, 2009.

Stech, Frank J., and C. Elässer. "Deception Detection by Analysis of Competing Hypothesis," *Proceedings of the 2005 International Conference on Intelligence Analysis*, 2005.

Stevens, Susan M., Courtney C. Dornburg, Stacey ML Hendrickson, and George S. Davidson. "Individual and Group Electronic Brainstorming In an Industrial Setting." In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 52, no. 5, pp. 493-497. SAGE Publications, 2008.

Stout, David. "Panel Says 'Dead Wrong' Data on Prewar Iraq Demands Overhaul." *The New York Times*, March 31, 2005.

Suedfeld, Peter Ed, and Philip E. Tetlock. *Psychology and Social Policy*. Hemisphere Publishing Corp, 1992.

Surowiecki, James. *The Wisdom of Crowds*. 2004. New York: Random House, 2004.

Team B Act. House of Representatives, 112th Congress (2011), HR 1502.

Tetlock, Philip. *Expert Political Judgment: How good is it? How can we know?*. Princeton University Press, 2005.

Tetlock, Philip E., and Barbara A. Mellers. "Intelligent Management of Intelligence Agencies: beyond accountability ping-pong." *American Psychologist*, Volume 66, Issue 6, 2011.

Tolcott, Martin A., F. Freeman Marvin, and Paul E. Lehner. "Expert decision-making in evolving situations." *Systems, Man and Cybernetics, IEEE Transactions*, Volume 19, Issue 3, 1989, pp. 606-615.

Toulmin, Stephen, *The Uses of Argument*, Cambridge: Cambridge University Press 1958.

Treverton, Gregory F., and C. Bryan Gabbard. *Assessing the Tradecraft of Intelligence Analysis*. Rand Corporation, 2008.

Treverton, Gregory F. *Intelligence for an Age of Terror*. New York: Cambridge University Press, 2009.

Tufte, E.R., *The Cognitive Style of PowerPoint*, Cheshire, CT: Graphics Press, 2003.

Thompson, Leigh and Elizabeth Ruth Wilson, "Rethinking the Wisdom of the Crowd: Why Individuals are More Creative than in their Groups," *The European Financial Review*, October 29, 2013, available at: <http://www.europeanfinancialreview.com/?p=716>

Trent, Stoney A., Emily S. Patterson, and David D. Woods. "Challenges for Cognition in Intelligence Analysis." *Journal of Cognitive Engineering and Decision Making*, Volume 1, Issue 1, 2007, pp. 75-97.

Treverton, Gregory F. *Addressing "Complexities" in Homeland Security*. Vol. 4. No. 973. The Swedish National Defence College, 2008.

Tung, Lai Lai, and Alan R. Heminger. "The Effects of Dialectical Inquiry, Devil's Advocacy, and Consensus Inquiry methods in a GSS environment." *Information and Management*, Volume 25, Issue 1, 1993, pp. 33-41.

Tung, Lai, and Mohammed Quaddus. "Conflict Management in Dialectical Inquiry, Devil's Advocacy and Consensus-Based Decision Making Approaches in a GSS Environment." *PACIS 2001 Proceedings*, Volume 11, 2001.

Tversky, Amos, and Daniel Kahneman. "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment." *Psychological review*, Volume 90, Issue 4, 1983.

Ulfelder, Jay, "Jay Ulfelder on the Rigor-Relevance Tradeoff" *The Good Judgment Project*, May, 2014, available at: <https://goodjudgmentproject.com/blog/?p=200>

U.S. Department of State, "Bureau of Intelligence and Research" no date, available at: <http://www.state.gov/s/inr/>

U.S. Government. *A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis*. Center for the Study of Intelligence Analysis, 2009. Available at:

[https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft% 20Primer-apr09. pdf](https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft%20Primer-apr09.pdf)

Walt, Stephen. "Rigor or Rigor Mortis: Rational Choice Theory and Security Studies." *International Security*, Volume 23, Number 4, Spring 1999, pp. 5-48.

Walt, Stephen, "Policy analysis in global affairs: What should my students read?" *ForeignPolicy.com*, November 2011, available at: http://www.foreignpolicy.com/posts/2011/11/22/policy_analysis_in_global_affairs_what_should_my_students_read

Warner, Michael. *Wanted: A Definition of Intelligence*. Washington, DC: The Center for the Study of Intelligence, 2002.

Wason, Peter C. "Reasoning about a rule." *The Quarterly Journal of Experimental Psychology* Volume 20, Issue 3, 1968, pp. 273-281.

Wastell, Colin A., et al. "The Impact of Closed-mindedness on the Assessment of Threat: An Empirical Study." *Open Psychology Journal*, Volume 6, Issue 1, 2013.

Weiss, Carol H., and Michael J. Bucuvalas. "Truth tests and utility tests: Decision-makers' frames of reference for social science research." *American Sociological Review*, 1980, pp. 302-313.

"Welcome to the CEBCP" *The Center for Evidence-Based Crime Policy*, 2013, accessed 26, 2015. Accessible at: <http://cebcp.org/>

"What is a Systematic Review?" *The Campbell Collaboration*, Accessed 21 April, 2015, available at: http://www.campbellcollaboration.org/what_is_a_systematic_review/

Wheaton, Kristan, "Reduce Bias In Analysis: Why Should We Care?" 25 March, 2014, available at: <http://sourcesandmethods.blogspot.com/2014/03/reduce-bias-in-analysis-why-should-we.html>

Wheaton, K. J., and D. E. Chido. "Structured analysis of competing hypotheses: improving a tested intelligence methodology." *Competitive Intelligence Magazine*, Volume 9, Issue 6, 2006

Wohlstetter, Roberta. *Pearl Harbor: Warning and Decision*. Stanford, CA: Stanford University Press, 1962.

Yin, Robert K., and Karen A. Heald. "Using the Case Survey Method to Analyze Policy Studies." *Administrative Science Quarterly*, 1975, pp. 371-381.

Zegart, Amy B. "Universities Must Not Ignore Intelligence Research." *Chronicle of Higher Education*, Volume 53, Issue 45, 2007.

Zelik, Daniel J., Emily S. Patterson, and David D. Woods. "Measuring Attributes of Rigor in Information Analysis." *Macroognition Metrics and Scenarios: Design and Evaluation for Real-World Teams*, 2010, pp. 65-83.

Ziegler, Rene, Michael Diehl, and Gavin Zijlstra. "Idea Production in Nominal and Virtual Groups: does computer-mediated communication improve group brainstorming?." *Group Processes and Intergroup Relations*, Volume 3, Issue 2, 2000, pp. 141-158.